

Deskription und Inferenz: methodologische Konzepte in der Statistik und Ökonometrie

Rahlf, Thomas

Veröffentlichungsversion / Published Version
Themenheft / topical issue

Empfohlene Zitierung / Suggested Citation:

Rahlf, T. (1998). Deskription und Inferenz: methodologische Konzepte in der Statistik und Ökonometrie. *Historical Social Research, Supplement*, 9, 1-319. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-286000>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

HSR

Supplement / Beiheft

No. 9 (1998)

Thomas Rahlf

Deskription und Inferenz
Methodologische Konzepte in der
Statistik und Ökonometrie

Köln
Zentrum für Historische Sozialforschung
1998

Inhaltsverzeichnis

Inhaltsverzeichnis	3
Vorwort	5
Einleitung	7
A. Konzepte der Statistik	16
1. Allgemeines	16
2. Deskription	21
a. Allgemeines	21
b. Anfänge der Statistik als Beschreibung	22
c. Der Fehlerbegriff und die Entwicklung des Kleinstquadratverfahrens	24
3. Inferenz	28
a. Allgemeines	28
b. Entstehung der Inferenzstatistik	30
c. Beginn der modernen Statistik: F. Y. Edgeworth und K. Pearson	55
d. Grundlegung der modernen Schätztheorie: R. A. Fisher	64
e. Grundlegung der "klassischen" Sicht: J. Neyman und E. S. Pearson	70
f. Inferenz für gegebene Daten und Hypothesenwahrscheinlichkeit	75
4. Aspekte der Modellbildung	85
a. Allgemeines	85
b. Der Faktor Zeit: 'Die ganze Welt ist ein stochastischer Prozeß'	88
c. Deskription und Inferenz in der explorativen Datenanalyse	108
d. Ausblick: Neuere Entwicklungen	117
5. Zusammenfassung und kritische Würdigung	121
B. Konzepte der Ökonometrie	133
1. Allgemeines	136
2. Deskription und Exploration	136
a. Allgemeines	136
b. Trends, Zyklen, Korrelationen	137
c. Die Rolle der Zufallsgrößen	153
d. Der traditionelle Weg	160
3. Inferenz und Modellbildung	167
a. Allgemeines	167
b. Anfänge ökonometrischer Traditionen	170
c. Noch einmal Konjunkturzyklen: Die Tinbergen-Keynes-Kontroverse	177
d. Übertragung der "klassischen" Sicht in die Ökonometrie	183
e. Ökonometrische Methodologien	196
4. Zusammenfassung und kritische Würdigung	222

Die Entwicklung der quantitativen Wirtschaftsgeschichte	235
Resümee: Ein Plädoyer für die Geschichte	256
Anhang	263
A1. Illustration des Likelihoodprinzips	263
A2. Illustration bayesianischer Inferenz	272
Abbildungs- und Tabellenverzeichnis	279
Literaturverzeichnis	281

Vorwort

Fast vier Jahrzehnte ist es nun her, da C. P. Snow die "zwei Kulturen" heraufbeschwor und damit heftige Diskussionen um das Verhältnis zwischen den Natur- und den sogenannten Geisteswissenschaften auslöste. Später wurden die Arbeiten von T. S. Kuhn und P. Feyerabend von manchem als Gegenangriff gesehen, und die Geisteswissenschaften konnten sich eines Triumphgefühls ob der offensichtlichen Relevanz historischer und soziologischer Komponenten auch im naturwissenschaftlichen Forschungsprozeß sowie der Relativierung naturwissenschaftlicher Wahrheitsansprüche nicht erwehren. Spätestens mit dem vielgerühmten Buch von H. Collins und T. Pinch "*The Golem: What everyone should know about science*"¹ schien eine solche Sichtweise auch unter den Naturwissenschaften allgemeine Akzeptanz gefunden zu haben.

Die Kontroverse ist jedoch aktueller denn je: der "science war" tobt. Nachdem P. Gross und P. Levitt 1994 insbesondere der akademischen Linken einen zunehmenden Antirealismus und -rationalismus vorwarfen, A. Sokal bei einer literaturwissenschaftlichen Zeitschrift einen – nicht ganz ernstgemeinten – Artikel über die "transformative Hermeneutik der Quantengravitation" einreichte und dieser zu allem Übel die Gutachter-Schranken passierte und gedruckt wurde, haben nicht nur Collins und Pinch sich einer breiten Front von Kritik ausgesetzt gesehen, die sogar von der Tagespresse rezipiert wurde.²

In dieser Situation ist es um so wichtiger, die Bedeutung wissenschaftshistorischer Ansätze herauszustellen. Zwar gibt es mittlerweile umfangreiche Studien über die Entwicklung der theoretischen Statistik und Ökonometrie, doch erst in jüngster Zeit macht sich auch in diesen "harten" Disziplinen ein Bewußtsein breit, daß sie nicht nur eine Geschichte haben, sondern diese auch viele Inkonsistenzen zu erklären vermag. An diese Entwicklung möchte die hier vorgelegte Arbeit aus Sicht eines Wirtschaftshistorikers anknüpfen und somit auch zu einer kritischeren Sicht bei der Anwendung solch unumstößlicher Wahrheiten, wie es die statistischen und ökonometrischen Methoden zu sein scheinen, beitragen. Schon der zeitliche und thematische Rahmen verdeutlicht, was die Arbeit nicht sein kann: eine umfassende Analyse der historischen Entwicklung dieser beiden Disziplinen bis zur Gegenwart. Sie sollte vielmehr als Leitfaden aufgefaßt werden, der – sicher in mehrfacher Hinsicht erweiterungs- und verbesserungsfähig – vieles kritischer sehen hilft und auch im Ergebnis der quantitativen Wirtschaftsgeschichte hoffentlich einen größeren Stellenwert zu geben vermag.

¹ Demnächst ergänzt durch: *Golem at work: What everybody should know about technology.*

² Siehe dazu Lüthy (1997).

Die vorliegende Arbeit basiert auf meiner im April 1997 abgeschlossenen Dissertation "Deskription und Inferenz. Entstehung und Entwicklung methodologischer Konzepte in der Statistik und Ökonometrie und ihre Verwendung in der quantitativen Wirtschaftsgeschichte". Sie wurde für die vorliegende Veröffentlichung vor allem um eine Übersicht über die Lehrbuchsituation in der quantitativen Geschichtswissenschaft und drei mathematische Anhänge gekürzt. (Die Sozialwissenschaften verfügen über wenige harte und unumstrittene Gesetze. Eines der unumstrittensten lautet aber sicher: jede Formel reduziert die Leserschaft um 50 Prozent.)

Mein Dank gilt an erster Stelle den Betreuern. Professor Helmut Thome hat vor langer Zeit mein Interesse für quantitative Methoden geweckt, hatte stets für Fragen ein offenes Ohr und ließ mir am Institut für Soziologie genügend Freiraum, um die dortige angenehme Arbeitsatmosphäre in konkrete Ergebnisse umzusetzen. Professor Peter Hertner unterstützte mich nach meinem Wechsel zum Institut für Geschichte 1995 ebenso großzügig und nach Kräften, so daß die Arbeit trotz aller beruflichen Verpflichtungen zügig abgeschlossen werden konnte. PD Dr. Rainer Metz (Köln), der wohl wie kaum ein zweiter im deutschen Sprachraum die Probleme der Anwendung ökonometrischer Verfahren in der Wirtschaftsgeschichte kennt, hat mehr als einmal sein Interesse von Einheitswurzeln auf die Wurzeln der Statistik und Ökonometrie gelenkt und mir dabei wichtige Anregungen geliefert. Teilergebnisse der vorliegenden Arbeit konnte ich auf der 35. Sitzung der Arbeitsgruppe Prognoseverfahren der Deutschen Gesellschaft für Operations Research e.V. in Augsburg sowie bei einem Vortrag am Seminar für Wirtschaftsgeschichte des Volkswirtschaftlichen Instituts der Universität München präsentieren. Allen Diskussionsteilnehmern, insbesondere Prof. P. Kischka, Prof. J. Komlos, Prof. C. Hillinger, Dr. U. Woitek, Dipl. Stat. M. Heintel und Dr. J. Baten, danke ich für ihre Anregungen. Für hilfreiche Hinweise danke ich weiterhin Prof. R. Tilly, Dr. T. Hanf, Dr. R. Petri, Dr. M. Seils und Dr. S. Nikolow.

Die Dissertation wurde 1997 mit dem Dorothea-Erxleben-Preis der Martin-Luther-Universität Halle-Wittenberg ausgezeichnet. Ich danke der Kommission für die Vergabe der Universitätspreise und dem Stifter des Preises, der Deutschen Bank AG, für diese ehrenvolle Auszeichnung.

Nicht zuletzt gebührt mein Dank dem Zentrum für Historische Sozialforschung, insbesondere Prof. Wilhelm Heinz Schröder für die Aufnahme der Arbeit in die Supplementreihe der Zeitschrift *Historical Social Research / Historische Sozialforschung*.

Halle, im Dezember 1997

Thomas Rahlf

Einleitung

Wenn man in unseren Tagen eine Ausgabe der führenden amerikanischen wirtschaftshistorischen Zeitschrift, des *Journal of Economic History*, aufschlägt, so wird bei aller Themenvielfalt eine Gemeinsamkeit auffallen: In der Regel wird man eine quantitative wirtschaftshistorische Untersuchung vorfinden, in der ein bestimmtes *Modell* untersucht wird; bei diesem Modell wird es sich aller Wahrscheinlichkeit nach um ein solches handeln, das *Variablen* anhand von *Parametern* in eine funktionale Beziehung zueinander setzt. Dieses Modell wird anhand konkreter historischer Daten einer Prüfung unterzogen worden sein, und seine Parameter werden sich als "signifikant" erwiesen haben und somit ein historisches Phänomen oder eine Entwicklung durch andere Phänomene *erklären*, wobei die Erklärungskraft vermutlich durch eine Zahl, einen *Determinationskoeffizienten*, wiedergegeben wurde.

Derartige Untersuchungsansätze prägen, gleich ob es sich hierbei um die Determinanten der industriellen Entwicklung, die kausalen Abhängigkeiten des Zins- und Preisniveaus im 18. oder 19. Jahrhundert in England, Frankreich oder den USA oder um die Einflußfaktoren des mittelalterlichen städtischen Getreideangebotes handelt, das Bild der amerikanischen und zunehmend auch der angelsächsischen und deutschen wirtschaftshistorischen Forschung.

Die Arbeiten dieser sogenannten "*new economic history*" belegen eine Tendenz, die von Robert Whaples als "*cliometric revolution*" bezeichnet wurde und 1993 mit der Verleihung des Nobelpreises für Wirtschaftswissenschaften an zwei "*Cliometriker*", Robert Fogel und Douglas North, auch internationale Beachtung und Anerkennung gefunden hat.¹ Whaples, der den Anteil "*cliometrischer*"² Artikel bzw. den Anteil entsprechender Seiten anhand des *Journal of Economic History* untersucht hat, stellte fest, daß dieser Anteil von 15.8 % zwischen 1961 und 1965 auf 80.5 % zwischen 1986 und 1990 gestiegen ist.³ Überblickt man die nach 1990 bis zur Gegenwart erschienenen Beiträge, so ist bereits ohne Auszählung ersichtlich, daß der Anteil seither nochmals gestiegen ist.

Überraschend ist in diesem Zusammenhang ein anderes Ergebnis von Whaples: Der Anteil methodologischer Beiträge/Seiten ist von 6.5 % 1941-50 auf 0.3 % 1981-90 gesunken.⁴ Dieses Phänomen ist unserer Ansicht nach nicht mit einer fortschreitenden Klärung methodologischer Fragen zu begründen, sondern mit einer zunehmend unkritischen Übernahme externer methodologischer Standards, namentlich aus der Ökonometrie.

¹ Whaples (1991).

² Er verwendete drei alternative Definitionen für cliometrische Artikel, denen allen die explizite Verwendung ökonomischer Theorie und Messung gemein ist (S. 293f). Die strengste Definition, die wir hier alleine wiedergeben, fordert die Verwendung einer Regressionsanalyse.

³ Vgl. Whaples (1991, S. 293, Tab. 3).

⁴ Vgl. ebda., S. 290, Tab. 1.

Seit den 1960er Jahren werden auch in der Wirtschaftshistoriographie, vor allem in der Konjunktur- und in der monetären Wirtschaftsgeschichte, in zunehmendem Umfang zeitreihenanalytische Verfahren und Modelle aus dem Bereich der Ökonometrie adaptiert. Während der Beginn dieser Phase zum einen durch eine eher unkritische Übernahme bestimmter Techniken, zum anderen durch eine Polarisierung der Diskussion gekennzeichnet war,⁵ hat in jüngster Zeit vereinzelt ein Nachdenken über die Angemessenheit der Anwendung verschiedener Verfahren eingesetzt. Diese Situation wurde nicht zuletzt durch die neuere Entwicklung in der Ökonometrie ausgelöst, die in zunehmendem Maße zeitreihenanalytische Modelle verwendet, um somit dem Zeitreihencharakter der von ihr untersuchten Daten stärker Rechnung zu tragen.⁶ Dort hat in den letzten Jahren ein methodisches Umdenken stattgefunden, das statt der Dynamisierung ursprünglich statisch konzipierter Theorien und Modelle nun immer stärker Zeitreihenmodelle in den Vordergrund stellt. Solch eine Situation ist einerseits zu begrüßen, da sie zu einer neuerlichen Annäherung von Ökonomie und Wirtschaftsgeschichte führen könnte, andererseits sind die Gefahren dieser Entwicklung nicht zu übersehen. Im Laufe der letzten hundert Jahre haben sich divergierende Erkenntnisziele in der Ökonomie und der Wirtschaftsgeschichte herausgebildet, die eigentlich in unterschiedlichen Verfahren und Modellen, zumindest aber in unterschiedlichen Interpretationen ihren Ausdruck finden müßten. Momentan sehen wir uns jedoch mit einer eigenartigen Gemengelage konfrontiert: In der Wirtschaftsgeschichte werden die quantitativen Verfahren in der Regel von der Ökonometrie adaptiert, ohne daß deren Bedeutungsgehalt bislang einer systematischen Analyse unterzogen worden wäre. Es erweist sich somit als sinnvoll und notwendig, diese Anwendung von Verfahren und Modellen in der quantitativen Wirtschaftsgeschichte, besonders aber die damit verbundenen *Konzepte*, näher zu überprüfen.⁷

Die folgenden Fragen sind unseres Erachtens in diesem Zusammenhang von besonderem Interesse. Wenn das Ziel einer statistischen Analyse nicht nur eine zusammenfassende Beschreibung einer historisch einmaligen Situation darstellen soll:

- Was bedeutet es, wenn etwa für makroökonomische Zusammenhänge ein Parameter als "signifikant" bezeichnet wird? Ist dies für wirtschaftshistorische Analysen eine sinnvolle Angabe über die Unsicherheit einer Schätzung?

⁵ Vgl. dazu beispielhaft die in Wehler (1973a, Teil 2 und 3) abgedruckten Beiträge.

⁶ Stellvertretend sei hier auf Hamilton (1994, S. XIII) verwiesen: "There has been an explosion of research in this area in the last decade, as 'time series econometrics' has practically come to be synonymous with 'empirical macroeconomics'." Wenn wir im folgenden von "der" Zeitreihenanalyse sprechen, verwenden wir dies als Sammelbezeichnung für jene Verfahren, die zur Analyse von Zeitreihen angewendet werden.

⁷ Unter "Konzept" verstehen wir Ideen, die Kategorien sowie Beschreibungen wichtiger Eigenschaften dieser Kategorien beinhalten.

- Eine “Signifikanzaussage” beinhaltet eine Aussage über Wahrscheinlichkeiten. Wie aber ist dieser Begriff überhaupt zu verstehen? Kann man ihn, wie dies üblicherweise geschieht, auch in einem wirtschaftshistorischen Zusammenhang als Grenzwert einer relativen Häufigkeit auffassen?
- Wirtschaftshistorische Aussagen beziehen sich oftmals nicht nur auf einen Zustand, sondern auf eine Entwicklung. In welcher Weise wird dem Aspekt der Zeit in diesem Zusammenhang Rechnung getragen?

In der Regel finden bei quantitativen Daten Modelle der Form

$$y = f(x_1, x_2, \dots) + \epsilon, \text{ konkret zumeist } y = \beta_1 x_1 + \beta_2 x_2 + \dots + \epsilon,$$

Verwendung. Auch diesbezüglich ließen sich eine Reihe von Fragen stellen, wie zum Beispiel:

- Wieso geht man von einer linearen Beziehung im Rahmen eines Regressionsmodells aus, dessen Parameter in der Regel mittels des Kleinstquadratverfahrens oder der Maximum-Likelihood-Methode bestimmt (“geschätzt”) werden? Warum verwendet man gerade diese Techniken? Die Bestimmung der Parameter durch Minimierung der quadrierten Abstände oder durch Maximierung der Likelihoodfunktion erfolgt aufgrund von Kriterien, die in einem spezifisch statistischen Sinne Optimalität beurteilen. Ist die Anwendung dieser Kriterien auch für historische Untersuchungen selbstverständlich?
- Die Parameter werden in der Regel als im betrachteten Zeitraum konstant angesehen. Resultiert diese stillschweigende Annahme aus historischen Erwägungen, die generell eine Strukturkonstanz im untersuchten Zeitraum plausibel erscheinen lassen, oder wird nicht eher von der Statistik und Ökonometrie ein Instrumentarium adaptiert, ohne über dessen Angemessenheit für die Erkenntnisziele der Wirtschaftsgeschichte zu reflektieren?
- In der Regel liegen keine unabänderlichen Modelle vor, über deren Richtigkeit keinerlei Zweifel besteht. Oftmals werden Modelle nach einer Inspektion der vorhandenen Daten modifiziert oder gar erst aufgestellt. Ist ein solches Vorgehen mit der Logik der verwendeten Verfahren vereinbar?

Nun kann eine Beantwortung derartiger Fragen nicht enzyklopädistisch für alle zeitreihenanalytischen Verfahren und Modelle, die in der quantitativen Wirtschaftsgeschichte angewendet werden, bewältigt werden. Ein erster, umfassender Versuch, eine Gruppe von Verfahren zu würdigen, wurde vor kurzem von Metz (1995) unternommen, der der Frage nachging, welche Bedeutung den von der Ökonometrie in den letzten Jahren intensiv diskutierten sogenannten “stochastischen Trends” für die Wirtschaftsgeschichte zukommt.⁸ Diese Arbeit macht unmißverständlich klar, worauf sich die Wirtschaftsgeschichte einlassen muß, will sie zu einer kritischen Beurteilung der neueren Entwicklungen in der Ökonometrie in der Lage sein und zugleich ihre Eigenständigkeit bewahren: Neben der Rezeption einer

nahezu unübersehbaren Fülle von Verfahren muß vor allem deren *Grundlage* erkannt werden. Wir wollen uns hier daher nicht einem weiteren Spezialgebiet widmen, sondern die Grundlagen selbst in den Mittelpunkt stellen.

Wie tief sollte eine solche Analyse aber gehen? Gemäß dem Diktum Max Webers, daß es "für logische Untersuchungen [...] schlechthin nichts 'Selbstverständliches'" gibt,⁹ sollte es das Ziel sein, das "Warum so und nicht anders" möglichst tief zu erfassen.¹⁰ Gerade durch die wachsende Anwendung stark formalisierter mathematischer Verfahren ist die Gefahr besonders groß, daß aufgrund der für den Wirtschaftshistoriker meist ungewohnten technischen Formelsprache Voraussetzungen, Bedingungen oder Annahmen nicht hinterfragt und auf ihren logischen Gehalt hin überprüft werden, weil die mathematische Stringenz auch eine logische Stringenz suggeriert.

Wie sollte ein solches Vorhaben realisiert werden? Wenn man versucht, eine Basis für die Entwicklung der methodologischen Konzepte (oder ein "Forschungsprogramm") zu schaffen, so gilt es zunächst, hierfür eine wissenschaftstheoretische Orientierung zu suchen. Orientiert man sich naheliegenderweise an den Sozialwissenschaften, so wird – zumindest für den quantitativen Bereich – zumeist eine mit Blick auf die Naturwissenschaften entwickelte Wissenschaftstheorie zugrundegelegt, die entweder Karl R. Poppers "Kritischem Realismus" oder der teilweise darauf aufbauenden "Methodologie von Forschungsprogrammen" von Imre Lakatos verpflichtet ist.¹¹

Wir wollen diesen Weg nicht einschlagen und statt dessen einen "historistischen" Ansatz wählen. Der Begriff des Historismus hat sich zahlreichen Bedeutungswandeln und Mißverständnissen ausgesetzt gesehen. Wir meinen damit weder teleologische Geschichtskonzeptionen eines Hegel oder Marx noch Deutungen, die das Einmalige als Kern jedweder geschichtlicher Entwicklung (F. Meinecke) betrachten, sondern: die Überzeugung, daß die Wirklichkeit (und somit auch eine Methodologie) 1. nur in ihrer historischen Entwicklung verstanden werden kann und 2. Paradigmen zeit- und raumbunden im (sozialen) Rahmen einer "scientific com-

⁸ Metz hat sich in einer Reihe früherer Arbeiten mit der Anwendung verschiedener Verfahren insbesondere in bezug auf die Untersuchung der sog. "langen Wellen" auseinandergesetzt. Vgl. unter anderem Metz (1984), Metz (1988a), Metz (1988b) oder Metz (1993).

⁹ Weber (1903-1906 [1968, S. 3]).

¹⁰ Andererseits muß auch bedacht werden, daß "[...] die Begründung der Methoden, die eine empirische Wissenschaft benutzt, irgendwann abgebrochen werden muß, damit sich kein unendlicher Regreß ergibt." Brinkmann (1970, S. 210).

¹¹ Popper (1994), Lakatos (1974). Für die Sozialwissenschaften siehe z. B. die kurze Übersicht in Schnell/Hill/Esler (1993, S. 103-115). Für die Ökonomie geben die Sammelbände von de Marchi (1988) und de Marchi/Blaug (1991) einen Überblick über den Stand der Diskussion. Starke Beachtung in der methodologischen Diskussion der Ökonomie haben auch die zum Teil auf P. Feyerabend aufbauenden Arbeiten von D. McCloskey gefunden, insbesondere McCloskey (1985a), McCloskey (1990) und McCloskey (1993). Sie sind in ihrer Art jedoch grundsätzlicher Natur und daher kaum auf quantitativ-methodologische Probleme anwendbar.

munity" entstehen, in der sich die Beteiligten über die Regeln der Forschung und des wissenschaftlichen Diskurses einig sind, diese wiederum mit anderen, von anderen Forschern oder Forschergruppen aufgestellten Paradigmen inkommensurabel sind und daß 3. als Konsequenz daraus der transdisziplinäre Transfer von Paradigmen nicht als kumulativer Fortschrittsprozeß angesehen werden kann, sondern eben wissenschaftshistorisch begründet werden muß.

Solch eine Vorgehensweise läßt sich zunächst auch mit K. R. Popper begründen, dessen Ausgangspunkt in *Logik der Forschung* darin bestand, "(...) herauszufinden, was andere über das vorliegende Problem gedacht und gesagt haben: warum es ein Problem für sie war; wie sie es formuliert haben; wie sie es zu lösen versucht haben. Das scheint mir ein wesentlicher Schritt in der allgemeinen Methode der rationalen Diskussion zu sein."¹² Damit enden jedoch auch die Gemeinsamkeiten. Die Analyse des vorliegenden Gegenstandes wird zeigen, daß sich in diesem Bereich Wissenschaften *nicht* als ein Prozeß realisiert haben, in dem Theorien falsifiziert und durch bessere ersetzt wurden, mithin die Wissenschaften einen kumulativen Fortschrittsprozeß beschreiben. Hier zeigt sich vielmehr ein Prozeß, wie er vor allem von Thomas S. Kuhn postuliert wurde.¹³ In diesem wissenschaftstheoretischen Ansatz kommt der "Geschichtsschreibung" eine zentrale Rolle zu: "Wenn man die Geschichtsschreibung für mehr als einen Hort von Anekdoten oder Chronologien hält, könnte sie eine entscheidende Verwandlung im Bild der Wissenschaft, wie es uns zur Zeit gefangen hält, bewirken."¹⁴ Eine Wissenschaft ist nach Kuhn nur aus ihrer historischen Entwicklung heraus zu verstehen. Die traditionelle Wissenschaftsvermittlung liefere dagegen ein ahistorisches Bild:

"Dieses Bild ist, sogar von Wissenschaftlern selbst, bisher in erster Linie nach dem Studium abgeschlossener wissenschaftlicher Leistungen gezeichnet worden, wie man sie bei den Klassikern und in neuerer Zeit in den Lehrbüchern für die junge wissenschaftliche Generation findet. Der Zweck solcher Bücher ist jedoch zwangsläufig der, zu überzeugen und pädagogisch zu wirken; eine aus ihnen gewonnene Konzeption der Wissenschaft paßt genausowenig auf ihre wirkliche Entstehung, wie es das Bild einer nationalen Kultur tun würde, das man aus einem Touristenführer oder einer Sprachlehre gewinnt."¹⁵

Definiert man den von Kuhn geprägten Begriff des Paradigmas¹⁶ als "mustergültige Problemlösungen, welche zum Vorbild der gesamten Forschung auf einem

¹² Popper (1994, S. XVf).

¹³ Siehe vor allem Kuhn (1976). Weitere Ansätze, die in die gleiche Richtung und zum Teil darüber hinausgehen (Kurt Hübner, Yehuda Elkana) sowie Vorläufer wurden von Scholtz (1991, S. 158-200) zusammengefaßt.

¹⁴ Kuhn (1976, S. 15). Für die Ökonometrie betonte Kirchgässner (1983, S. 531) die Bedeutung dieses Ansatzes.

¹⁵ Kuhn (1976, S. 15).

¹⁶ Wir wollen den zahlreichen Fehlinterpretationen und Überstrapazierungen des Paradigmenbegriffes seit seiner Verwendung durch Kuhn nicht nachgehen. Gerade in den Sozial- und Geisteswissenschaften hat dieser mit Blick auf die Physik verwendete Begriff nach seiner ersten Verbreitung in den sechziger und siebziger Jahren einen schlagwortartigen Charakter erlangt. Vgl. etwa für die Geschichtswissenschaft die in Kocka (1986, S. 240, Anm. 112a)

bestimmten Fachgebiet werden”,¹⁷ so ergibt sich 1. eine natürliche Ordnung der vorhandenen methodologischen Konzepte und 2. eine Begrifflichkeit, mit der vorhandene Widersprüche besser erklärt (oder verstanden) werden können als mit logisch-theoretischen Kategorien, die Forschung als linearen und kumulativen Prozeß begreifen.

Um Mißverständnissen vorzubeugen: Wir wollen hier nicht die Tragfähigkeit der Kuhnschen Position überprüfen, sondern sie lediglich als Leitidee verwenden, um einen Weg für methodologische Konzepte für die Wirtschaftsgeschichte zu bahnen. Es ist nicht unser Ziel, die Kuhnsche Theorie zu beurteilen.¹⁸

Kommen wir nun zur konkreten Ausführung der Konzeption. Aus Gründen der analytischen Klarheit muß natürlich auch eine Darstellung einer historischen Entwicklung systematisch gegliedert werden. Die systematische Gliederung ist hier zugleich auch eine hierarchische: Die quantitative Wirtschaftsgeschichte orientiert sich nicht erst seit der “cliometrischen Revolution” an Methoden und Konzepten der Wirtschaftswissenschaften insbesondere der Ökonometrie. Diese wiederum hat jedoch ihrerseits Paradigmen der theoretischen Statistik übernommen, die losgelöst von den spezifischen Problemen und Erkenntniszielen der Ökonom(etr)ie betrachtet werden müssen. Innerhalb dieser systematischen Kategorien wollen wir weitgehend chronologisch untergliedern.

Eine immer wieder anzutreffende Einteilung der *Statistik* unterscheidet zwischen Deskription und Inferenz.¹⁹ Die moderne Statistik begann als *Beschreibung*. Eine zentrale Rolle spielte hier zum einen die Entwicklung des Fehlerbegriffes, zum anderen – und eng damit zusammenhängend – die Entwicklung der Ausgleichsrechnung, speziell des Kleinstquadratverfahrens.

Die *Inferenzstatistik* hat sich erst im 20. Jahrhundert als eigene Disziplin selbständig ausgebildet. Bis zu diesem Zeitpunkt sind zentrale Konzepte in verschiedenen *Kontexten* entwickelt und übernommen worden.²⁰ Der weitere Verlauf ist durch drei “Grundlegungen” gekennzeichnet: erstens durch die Fundierung der modernen Statistik durch F. Y. Edgeworth und K. Pearson, zweitens die Grundlegung der modernen Schätztheorie durch R. A. Fisher und drittens schließlich die Grundlegung der “klassischen” Inferenzsicht durch J. Neyman und E. S. Pearson.

Parallel zu dieser Entfaltung der zentralen statistischen Konzepte hat sich eine eher philosophisch ausgerichtete Wahrscheinlichkeitstheorie weiterentwickelt und In-

angegebene Literatur. Die Unschärfe ist jedoch bereits bei Kuhn selbst vorhanden, der den Begriff des Paradigmas in mehreren Bedeutungen verwendet. Kuhn sah sich daher veranlaßt, den Begriff in einem 1969 erschienenen Postskriptum zu präzisieren.

¹⁷ Demmerling (1995, S. 477).

¹⁸ Vgl. dazu Hoyningen-Huene (1989) sowie Horwich (1993), wo sich auch neuere Antworten Kuhns auf seine Kritiker finden.

¹⁹ Nicht zuletzt ist dies die gängige Einteilung der Lehrbücher oder Lehrangebote.

²⁰ Porter (1987, S. 369): “Statistical reasoning evolved in the context of specific problems from the various disciplines, and grew through application to new subject matters, analysis from new viewpoints, and recognition of analogies among widely disparate fields.”

ferenzkonzeptionen nicht unbeeinflusst gelassen. Diese Parallelentwicklung gilt es an entsprechender Stelle jeweils zu berücksichtigen.

Während sich die von Neyman und Pearson begründete Inferenztheorie seit den dreißiger Jahren als dominierende Sicht durchsetzte, entstanden auch alternative Inferenzkonzeptionen: Dies sind zum einen verschiedene, sich selbst als "bayesianisch" bezeichnende Ansätze, die sich bis in die dreißiger Jahre unseres Jahrhunderts zurückverfolgen lassen, zum anderen das sogenannte Likelihood-Prinzip. *Logisch* ist eine Behandlung des Likelihood-Prinzips der bayesianischen Inferenz vorgeordnet. *Historisch* ist die Likelihood-Inferenz jedoch erst zu einem allgemeingültigen "Prinzip" erhoben worden, als die Grundlage der modernen bayesianischen *Wahrscheinlichkeitstheorie* bereits gelegt worden war. Es erweist sich auch hier als zweckmäßig, die historische Reihenfolge beizubehalten, da das Likelihood-Prinzip alleine nicht sehr weitgehende Inferenzaussagen ermöglicht und die meisten Verfechter dieses Prinzips es ohnehin mit einem bayesianischen Ansatz verbinden.

Sowohl Deskription als auch Inferenz bewegen sich in einem Rahmen, der sehr allgemein als "Modell" bezeichnet werden kann. Eine Deskription ist nicht ohne "Modellannahmen" möglich, und Inferenzaussagen sind vollständig von dem a priori zugrundegelegten Modell abhängig. Da in den wenigsten Situationen das Modell stets als a priori gegeben angenommen werden kann, ist dem Aspekt der Modellbildung bzw. -konstruktion Beachtung zu schenken. Dies geschieht zunächst in zweierlei Hinsicht.

Zum einen wollen wir der Berücksichtigung der Dimension "Zeit" im Rahmen der statistischen Modellierung unser Interesse zuwenden, zum anderen die sogenannte Explorative Datenanalyse näher behandeln. In diesem – keineswegs einheitlichen – Rahmen sind Ansätze entwickelt worden, die dem verbreiteten *iterativen* Vorgehen Rechnung tragen sollen. Schließlich wollen wir noch auf einige neuere Entwicklungen der allgemeinen Statistik hinweisen. Hier wurde in letzter Zeit insbesondere der Problematik inferentieller Aussagen, die nicht aufgrund eines unabhängigen, a priori existierenden Modells gewonnen wurden, nachgegangen. Vor diesem Hintergrund, der in einer kritischen Würdigung zusammengefaßt wird, können wir anschließend zentrale Konzepte der Ökonometrie behandeln.²¹ Hier ist die Trennung der drei Aspekte Deskription, Inferenz und Modellbildung in stärkerem Maße als im vorangehenden Teil auch eine chronologische: Anfangs wurde die Anwendung quantitativer Methoden in der Ökonomie in einem deskriptiven und "explorativen" Sinne betrieben. Naturgemäß kommt im sozioökonomischen

²¹ Es ist allerdings zu betonen, daß insbesondere in der frühen Phase eine Trennung zwischen Konzepten der Statistik und solchen der Ökonometrie nicht eindeutig möglich ist. Wir wollen hier folgende "Daumenregel" definieren: I) Wenn eine neue Methode eingeführt wird, die allgemeinen Anspruch erhebt, also nicht auf spezifisch ökonomische Probleme zugeschnitten ist, so wird diese Entwicklung in Teil A behandelt. II) Wenn eine Methode aus der in I) definierten Entwicklung der Statistik adaptiert wurde oder selbständig Methoden entwickelt wurden, die aber als spezifisch auf die Probleme der Ökonomie zugeschnitten angesehen wurden, so wird dies in Teil B behandelt.

Kontext der Beschreibung konkreter, d. h. einmaliger Entitäten eine größere Bedeutung zu, als etwa in der Physik oder der Biologie.²² Das Hauptaugenmerk galt in diesem Kontext Zeitreihen, wobei neben der Frage nach Zusammenhängen zwischen ökonomischen Entwicklungen vor allem die Analyse von Trends und Zyklen im Mittelpunkt des Interesses stand.

Von ungleich größerer Bedeutung für die Analyse ökonomischer Daten war die Entwicklung der "Ökonometrie", die 1929 in der Gründung der *Econometric Society* ihren förmlichen Ausdruck fand. Zu Beginn vor allem von R. Frisch, T. C. Koopmans und J. Tinbergen vorangetrieben, gab erst T. Haavelmo der ökonometrischen Forschung eine auf der Neyman-Pearson-Theorie aufbauende inferenzstatistische Basis, die auch die Ökonometrie bis heute dominiert.

In der Ökonometrie stellt sich das Problem der Zeit in besonderem Maße; nicht etwa, weil der Zeit in der ökonomischen Theorie eine besondere Rolle zukommt, sondern aus dem einfachen Grund, weil makroökonomische Daten in der Regel lediglich in Form von Zeitreihen vorliegen und damit Eigenschaften aufweisen, die eine Anwendung der herkömmlichen Verfahren in "Schätz"situationen als fragwürdig erscheinen lassen. Die in den dreißiger Jahren bereits vorherrschende Ahistorisierung der Ökonomie hat es konsequenterweise mit sich gebracht, daß die Berücksichtigung von Zeit nicht als Erkenntnisgewinn, sondern eher als lästiges technisches Problem angesehen wurde. Zwar hatte G. U. Yule bereits in den zwanziger Jahren bahnbrechende Arbeiten zum Problem statistischer Analysen von Zeitreihen vorgelegt, doch hatten diese eher negative Auswirkungen auf die Anfänge der empirischen Wirtschaftsforschung: "In the field of economic research the application of relations of pure theory to time series or historic records has become something like taboo. Many economists, not sufficiently trained in statistical theory, have, it seems, 'scared away' by such critical work as, e.g., that of G. U. Yule."²³ Zu dieser Zeit war man allgemein der Auffassung, daß inferenzstatistische Aussagen über Daten, die in Form von Zeitreihen vorliegen, nicht möglich seien. So schrieb etwa Schumpeter 1939 in seinem monumentalen Werk *Business Cycles*: "Das einzige, was ganz allgemein von Zeitreihen gilt, ist, daß sie die Wahrscheinlichkeitserfordernisse nicht erfüllen."²⁴ Die angemessene Berücksichtigung des Faktors Zeit sollte sich als zentral für die gesamte Entwicklung der Ökonometrie erweisen.

Der letzte Abschnitt dieses Kapitels stellt verschiedene ökonometrische "Methodologien" gegenüber. Die Konzeption T. Haavelmos wurde zu Beginn der fünfziger Jahre von der einflußreichen *Cowles Commission* übernommen und entwickelte sich für Jahrzehnte zur allgemein anerkannten Sichtweise in der Ökonometrie. Obwohl bereits früher Alternativen vorlagen, haben sich erst seit Ende der siebziger Jahre drei Methodologien herausgebildet, die entweder an solche Alternativen

²² Auch in der Physik oder Biologie werden Atome oder Gattungen *beschrieben*, aber wohl kaum *historisch einmalige* Realisationen dieser Atome oder Gattungen.

²³ Haavelmo (1944, S. 17). Er verweist auf Yule (1926).

²⁴ Schumpeter (1939 [1961, S. 207]).

anknüpften (C. Sims), völlig andere (bayesianische) Wege gingen (E. E. Leamer), den Modellbildungsprozeß explizit in klassischem Rahmen einbetteten (D. Hendry) oder eher pragmatisch argumentierten (M. Friedman).

Auch diesen Teil wollen wir in einer kritischen Würdigung noch einmal zusammenfassen.

Das letzte Kapitel schließlich ist vor diesem Hintergrund der Verwendung methodologischer Konzepte in der quantitativen Wirtschaftsgeschichte gewidmet. Zum Schluß sollen einige Vorschläge dargelegt werden, in welche Richtung zukünftig methodologische Entwicklungen in der quantitativen Wirtschaftsgeschichte gehen sollten.²⁵

Die – ebenso schwierige wie reizvolle – Aufgabe, die für die quantitative Wirtschaftsgeschichte als fruchtbar empfohlenen *Konzepte* mit konkreten *Techniken* und inhaltlichen Fragestellungen zu verbinden, muß späteren Arbeiten vorbehalten bleiben, für die die vorliegende Untersuchung – hoffentlich – eine erste Grundlage bilden kann.²⁶

Speziell im Hinblick auf die Statistik, aber auch bezüglich der Ökonometrie schließen wir uns einer Einschätzung Stegmüllers an, die uns für die Behandlung des vorliegenden Themas unerläßlich erscheint. So wie er als Philosoph und Wissenschaftstheoretiker müssen wir als Historiker die Grenzen unserer Fähigkeit, die Relevanz eines fachfremden Stoffes zu bewältigen, von vornherein klarstellen:

“*Bescheidenheit* [...] hat sich auf die statistische Fachliteratur zu richten. Was hier von Spezialisten geleistet worden ist – leider in einer ‘dem gewöhnlichen Sterblichen’ kaum oder nur schwer verständlichen Sprache –, muß zunächst *verstanden*, d. h. begrifflich durchdrungen und auf seine Voraussetzungen und Konsequenzen hin analysiert werden. ‘Besser verstehen!’ und nicht ‘besser machen!’ muß die Devise beim Beginn der Arbeit lauten.”²⁷

²⁵ Seit Abschluß des Manuskriptes (April 1997) sind vor allem zwei Arbeiten erschienen, auf die hinzuweisen ist: Die Entstehung der Wahrscheinlichkeitstheorie wurde jetzt in einer Arbeit von Hauser (1997) untersucht. Sie schildert ausführlich die in Abschnitt A 3 b zusammengefaßte Entwicklung, endet jedoch vor der problematischen Zeit: vor Laplace und Bayes. Die Geschichte der Zeitreihenanalyse wird nun ausführlich von Klein (1997) behandelt. Diese Arbeit war uns leider nicht mehr zugänglich.

²⁶ Ein aus Sicht der Wirtschaftsgeschichte relevanter technischer Aspekt – die Modellierung stochastischer Zyklen – wurde von uns in Rahlf (1996c) behandelt.

²⁷ Stegmüller (1973, S. 1). Hervorhebungen im Original.

A. Konzepte der Statistik

1. Allgemeines

Unabhängig von den konkreten Erkenntniszielen einer Fachwissenschaft existiert eine allgemeine Basis der Statistik. Zwar hat sich die Statistik im Laufe der Zeit immer wieder durch Anstöße aufgrund fachspezifischer Probleme weiterentwickelt, doch haben diese Anstöße bis heute nur in geringem Umfang zu “fachspezifischen Statistiken” geführt. Vielmehr wurden solche Anstöße in ein allgemeines Lehrgebäude der Statistik eingebaut und hatten, davon ausgehend, Rückwirkungen in andere Wissenschaften.¹ Der Großteil statistischer Erkenntnis ist, diese These sei bereits hier gewagt, von fachspezifischen Problemen unabhängig. Sie ist vielmehr logischer und damit philosophischer Natur.² Die in der Biologie und Agrarwissenschaft entwickelten Konzepte – man denke nur an Namen wie K. Pearson, R. A. Fisher oder G. U. Yule – bilden ohne Zweifel auch die *logischen* Grundlagen für die derzeitige Anwendung der Statistik in den Sozialwissenschaften.

Das Gebäude der theoretischen Statistik ist jedoch keineswegs so geschlossen und stringent, wie es etwa in nahezu allen Einführungslehrbüchern in “die Statistik” vermittelt wird. Unbestritten dürfte eine grundsätzliche Einteilung in die beiden Erkenntnisziele *Deskription* und *Inferenz* sein. Statistik als Wissenschaft war ursprünglich, spätestens jedoch seit der sog. “Universitätsstatistik”, definiert als *Staatsbeschreibung*. Zwar gibt es wohl kaum einen anderen Begriff, der im Laufe der Jahrhunderte so fundamentalen Bedeutungswandel unterlag,³ doch hat der Aspekt der Beschreibung bis heute seinen festen, wenn auch vergleichsweise unbedeutenden Platz in der Statistik behalten. Eine oftmals mißverständene Interpretation erfährt in diesem Zusammenhang das für viele empirische Untersuchungen zentrale “Kleinstquadratverfahren”, das keineswegs notwendigerweise mit inferenzstatistischen Implikationen verbunden werden muß und anfangs auch nicht damit verbunden wurde.⁴

¹ Vgl. dazu auch Rahlf (1996a).

² Siehe als Beispiel etwa die von Cassel/Orear (1971) problematisierten Erkenntnisziele der Statistik in der Physik. Kempthorne (1971, S. 286) bemerkt dazu in einem Kommentar “[...] that the problems discussed and the questions raised are of the same *logical* nature as those which have arisen in biology and agriculture over the past seventy years. I say this not to imply in any way that physics is backward, but to give evidence for the view that the basic problems considered in the *discipline* of statistics are universal.” Hervorhebungen im Original.

³ Menges (1972, S. 9): “Es ist eine historische Kuriosität, daß der Name Statistik von einem Extrem, der Universitätsstatistik, über die – eine Mittelstellung einnehmende – Politische Arithmetik zum anderen Extrem, der angewandten Wahrscheinlichkeitslehre, gewandert ist (obgleich freilich bis heute die materielle Statistik ihn in einem praktischen Sinn benutzt).”

⁴ Die Ursprünge dieses Verfahrens und seine nachträgliche Einbettung in die Wahrscheinlichkeitstheorie im 19. Jahrhundert werden in einer grundlegenden Arbeit von Stigler (1986) analysiert. Menges (1972, S. 315) betont zu Recht, “[...] daß die Methode der kleinsten

Besonders deutlich zeigt sich diese Tatsache daran, daß Laplace, der seit etwa 1770 auf beiden Gebieten Entscheidendes zur Entwicklung beigetragen hatte, diese zunächst unabhängig voneinander behandelte und erst 1809/10 durch Gauss und Laplace eine Synthese zustande kam. Auch danach gab es Positionen, die eine Anwendung wahrscheinlichkeitstheoretischer Erwägungen im Zusammenhang mit dem Kleinstquadratverfahren explizit ablehnten.⁵

Der sich anschließende Abschnitt thematisiert die zweite Säule der Statistik, die statistische Inferenz. So unbestritten die obengenannte Unterteilung ist, wird auf der anderen Seite kaum ein anderer Bereich der modernen Wissenschaft so kontrovers diskutiert, wie die (erkenntnis-)theoretischen Grundlagen der Inferenzstatistik. Statistische Inferenz als “Überwindung der Ungewißheit durch induktive Schlüsse, die ihre Basis in empirischen Beobachtungen haben”,⁶ basiert auf einem zentralen wissenschaftstheoretischen Problem: dem Problem der Induktion.⁷ Statistisch-induktive Hypothesen weisen eine Eigenschaft auf, die ihre wissenschaftstheoretische Behandlung im Vergleich zu deterministischen Hypothesen ungleich schwieriger gestaltet; sie sind weder verifizierbar noch falsifizierbar. Damit sind zwei von K. R. Popper entwickelte und allgemein anerkannte wissenschaftstheoretische Grundpositionen, das Falsifikationskriterium⁸ und die Trennung von empirischen und metaphysischen Aussagen, auf statistische Hypothesen nicht anwendbar.⁹ Eine statistische Hypothese, die aufgrund gegebener Daten vom Umfang n nach einem bestimmten Kriterium abgelehnt wird, kann unter Umständen aufgrund neu hinzugenommener Daten $n+m$ anhand desselben Kriteriums nicht abgelehnt werden, obwohl die ersten n Daten identisch geblieben sind.¹⁰

Quadrate, obgleich sie entscheidungstheoretisch wie inferenztheoretisch interpretierbar ist, im Grunde gar kein Schätzverfahren im eigentlichen Sinn ist, sondern – so könnte man sagen – ein mathematisches Ausgleichsverfahren, das unter gewissen Aspekten Eigenschaften und Zweckbestimmungen eines Schätzverfahrens annimmt.”

⁵ Siehe Porter (1986, S. 10, Anm. 5).

⁶ Menges (1972, S. 266). Die *Encyclopedia of statistical sciences* enthält nicht weniger als drei Einträge unter dem Stichwort “Inference”. Vgl. Koch/Gillings (1983), Dawid (1983), Fraser (1983).

⁷ Vgl. zum Problem der Induktion die zentralen Texte von Popper (1994), Carnap/Stegmüller (1959), Stegmüller (1971) sowie Stegmüller (1973), dessen einleitendes Kapitel bezeichnenderweise “Jenseits von Popper und Carnap” überschrieben ist. Popper war der Ansicht, das Problem bereits 1927 gelöst zu haben. Vgl. Popper (1993, S. 1).

⁸ Auch Stegmüller (1973, S. 3) vertrat die Ansicht, “[...] daß die Poppersche Theorie von vornherein auf deterministische Hypothesen zugeschnitten ist.” Hervorhebung im Original. Der Standpunkt einer als “deduktiv-nomologisch” bezeichneten idealen Vorgehensweise, der von Best/Schröder (1988) sogar für die quantitative Geschichtswissenschaft gefordert wird (s. o.), ist lediglich für deterministische Hypothesen relevant, während er, wie Menges (1976, S. 296) betonte, “für die Statistik wenig hergibt”. Vgl. auch Ströker (1992, S. 29, Anm. 12).

⁹ Stegmüller (1973, S. 38): “Nur nebenher sei erwähnt, daß wegen der Nichtherleitbarkeit (im strengen Sinn) von Existenzbehauptungen aus Wahrscheinlichkeitsansätzen Poppers Versuch der Abgrenzung von empirischen Sätzen gegenüber metaphysischen Aussagen für den Fall statistischer Hypothesen zusammenbricht.” Vgl. dagegen wiederum Popper (1994, S. 207ff, Anhang *IV [1981], Anhang *XVIII [1982] und Anhang *XIX [1983]).

¹⁰ Vgl. Stegmüller (1973, S. 3, 54).

Die zentrale Frage, die sich hierbei stellt, ist, auf welche Weise von Daten auf Gesetzmäßigkeiten in Form von Parametern, Verteilungen und Modellen geschlossen werden kann. Eine angemessene Übersicht über diesen Komplex kann nur ansatzweise gelingen, denn dieses Gebiet ist anscheinend von unüberwindbaren Gegensätzen gekennzeichnet, wie etwa Stegmüller (1973, S. 1) konstatiert: "Es besteht bis zum heutigen Tag eine ungeheure Kluft zwischen logischen und wissenschaftstheoretischen Analysen von Begriffen der Prüfung, der Bestätigung und der Bewährung von Hypothesen auf der einen Seite und von Fachleuten im Gebiet der mathematischen Statistik angestellten Untersuchungen über diese Themenkreise auf der anderen Seite."¹¹ Falls sich jene logischen oder wissenschaftstheoretischen Aussagen finden, so sind diese nach Stegmüller teilweise mit der Situation in der Philosophie zu vergleichen:

"In mindestens zwei Hinsichten ähnelt die Statistik der Philosophie. Erstens darin, daß es zahlreiche einander bekämpfende Richtungen gibt, welche die gegnerischen für Stumpfsinn erklären. Zweitens darin, daß in beiden Bereichen eine starke Tendenz zum Denken in Schablonen besteht."¹²

Eine erste Orientierung bietet der Artikel *Logic of Statistical Reasoning* der *Encyclopedia of Statistical Sciences*.¹³ Nach Kyburg ist das statistische Denken bis in die jüngsten Tage zum großen Teil informell und intuitiv geblieben, unabhängig von der spezifischen Folge des Denkens: ob dies eine Handlung, das Akzeptieren einer Behauptung, die Ablehnung einer Hypothese, die Zuweisung einer Wahrscheinlichkeit sei oder andere Bereiche betreffe. Seine Ausführungen beschränken die Logik des statistischen Denkens jedoch auf die *Inferenzstatistik* und räumen einer Deskription keine eigenständige Rolle ein.

Ein besseres Verständnis wird unseres Erachtens erreicht, wenn wir zunächst einen Überblick geben, wie die zentralen Konzepte der Statistik entstanden sind. Die Statistik ist nicht als mathematische Disziplin gewachsen, sondern in einem komplexen Abhängigkeitsverhältnis verschiedenster Bereiche.¹⁴

Vor der Begründung der modernen Statistik durch K. Pearson war die Entwicklung durch folgende Elemente gekennzeichnet: Zwei zentrale Bereiche, die Kombination von Beobachtungen, die später zur Kleinstquadraterechnung führt und die

¹¹ In diesem Sinne auch Menges (1972, S. 24): "Noch immer ist die statistische Wissenschaft ohne einhellige Auffassung, selbst ohne *einigermaßen* einhellige Auffassung über ihre epistemologischen Grundlagen." Hervorhebung im Original.

¹² Stegmüller (1973, S. 76).

¹³ Kyburg (1985).

¹⁴ "Mathematics and physics may, as the Comtean hierarchy of the sciences suggests, have logical primacy over biological and social science, but historically the situation is much more complex and interesting." Porter (1986, S. 11). Wir werden nur diejenigen berücksichtigen, die für die heutigen Konzepte von Bedeutung sind und keinen umfassenden historischen Überblick über die Entwicklung der Statistik bis 1900 geben. Daher werden wir z. B. die Entwicklung der kinetischen Gastheorie durch Maxwell und Boltzmann, die durchaus einen wichtigen Abschnitt in der Geschichte der Statistik markiert, nicht behandeln. Vgl. dazu Porter (1986, S. 111-128). Lexis, dessen Ausführungen seinerzeit stark beachtet wurden, hatte keinen Einfluß auf die Entwicklung der allgemeinen Statistik, sondern auf eine erste speziell sozialwissenschaftliche, die unter dem Namen *kontinentale Schule* bekannt wurde.

Wahrscheinlichkeitstheorie entwickelten sich zunächst unabhängig voneinander. Der Beginn der Inferenzstatistik setzte mit Versuchen Jacob Bernoullis und Abraham De Moivres ein, nicht nur Wahrscheinlichkeiten aus Wahrscheinlichkeiten abzuleiten, sondern anhand von empirischen Beobachtungen unbekannte Wahrscheinlichkeiten zu bestimmen und die damit verbundene Unsicherheit quantitativ einzugrenzen. Ihnen gelang es jedoch nicht, den entscheidenden Schritt zu einer inversen Wahrscheinlichkeit zu vollziehen. Dieser konnte erst auf dem konzeptionell weniger problematischen Gebiet der Fehlerrechnung von P. S. Laplace vollzogen werden und gelang ihm und zur gleichen Zeit unabhängig davon auch T. Bayes durch eine Inversion des Wahrscheinlichkeitsbegriffes.

Die weitere Entwicklung ist durch einen Transfer dieser Konzepte in den sozialwissenschaftlichen Bereich gekennzeichnet. Laplace hatte bereits erste Beispiele untersucht, die dann von seinem Nachfolger S.-D. Poisson zu einem universalen "Anwendungsimperialismus" ausgebaut wurden und schließlich mit A. Quetelet zu einer "physique sociale" führten. Zwar handelt es sich hierbei um eine im nachhinein gesehen negative Entwicklung für die Anwendung statistischer Methoden in den Sozialwissenschaften, doch war dieser Weg für die Statistik als eigenständige Wissenschaft konstituierend, denn Quetelets Ausführungen waren es, die die Aufmerksamkeit von J. C. Maxwell auf das Fehlergesetz lenkten und ihn schließlich zusammen mit L. Boltzman zur Entwicklung der statistischen Gastheorie führten, die wiederum eine eigenständige Entwicklung in der Statistik hervorbrachte und schließlich in die Quantentheorie mündete.¹⁵ Auf die für unseren Zusammenhang relevanten Aspekte dieser und der weiteren Entwicklung wollen wir in den folgenden Abschnitten eingehen. Diese Abschnitte sollen Ansichten zu den Grundlagen der statistischen Inferenz klären.

Weiterhin muß jedoch ein dritter Aspekt behandelt werden, der in diesem Rahmen kaum Beachtung findet und in der Regel in die Rubrik "Voraussetzungen" abgeschoben wird: die *Hypothesenbildung* bzw. Aspekte der Modellkonstruktion. Beschränkt sich beim Aspekt der *Hypothesenprüfung* und Inferenz die Kontroverse immerhin noch auf die *Art und Weise*, stellt sich hinsichtlich des dritten Aspekts (*Hypothesenbildung*) die Meinungsverschiedenheit bereits bei dem "*Überhaupt*" ein. Eine Diskussion der Probleme, die mit der Inferenzstatistik, speziell mit der Hypothesenprüfung verbunden sind, setzt voraus, daß die Hypothesen als solche bereits vorhanden sind. Leider erweist sich die A-priori-Existenz derartiger Hypothesen oder "Modelle"¹⁶, vor allem in manchen Bereichen der Sozialwissenschaften, als unrealistisch. Ein Modell wird nicht immer unabhängig von den Daten zustande kommen.

Es muß die Frage gestellt werden, welche erkenntnistheoretischen Konsequenzen damit verbunden sind, wenn man aufgrund der Daten nicht eine vorgegebene Hypothese testet, sondern die Zulässigkeit der Hypothese selbst in Frage stellt, d. h. die Rahmenbedingungen, das "Modell" oder, wie Stegmüller es formuliert, die

¹⁵ Vgl. Porter (1986, S. 7).

¹⁶ Stegmüller (1973) nennt solche Hypothesen "Oberhypothesen".

“Oberhypothesen”. Und noch einen Schritt weiter: Lassen sich überhaupt Hypothesen aufgrund gegebener Daten aufstellen?

Für unseren Zusammenhang spielt hier der Faktor “Zeit” eine besondere Rolle, denn eine Reihe von Erkenntniszielen der quantitativen Wirtschaftsgeschichte ist auf die Analyse von Verläufen bzw. Prozessen ausgerichtet. Gerade die sehr formale und weit ausgebaute Theorie stochastischer Prozesse ist in einem Konglomerat unterschiedlichster Zusammenhänge und Spezialdisziplinen entstanden. Auch hier kann unseres Erachtens eine historische Betrachtungsweise aufschlußreich sein. In den dreißiger Jahren ist die grundlegende Theorie stochastischer Prozesse in ihren Prinzipien weitgehend abgeschlossen. Die weitere Entwicklung, die wir beschreiben, ist daher eher dem “angewandten” Aspekt gewidmet. Insbesondere das Problem der Prozeßbestimmung anhand konkreter Daten wurde zu einem zentralen Problem.

Aus den Problemen der Modellkonstruktion heraus ist in den letzten Jahren noch eine dritte Säule der Statistik hinzugekommen, über deren Stellung allerdings ebenfalls kontroverse Diskussionen entstanden sind: die *Exploration* oder *explorative Datenanalyse*. Zumeist mit einer Erweiterung der deskriptiven Statistik verbunden bzw. mit einer theoriefreien Vorgehensweise, ist sie nichts weniger als dieses oder jenes. Vielmehr wird man diesem Ansatz gerecht, wenn man ihn als Synthese der beiden obengenannten Erkenntnisziele der Statistik betrachtet. Vor allem J. W. Tukey, der gelegentlich als “Erfinder” der explorativen Datenanalyse (“EDA”) bezeichnet wird – zumindest stammt die Bezeichnung von ihm –, hat in zahlreichen Beiträgen auf diesen Charakter der Exploration hingewiesen.

In gewisser Hinsicht alternativ zu J. W. Tukey hat G. E. P. Box eine Datenanalysekonzepktion entwickelt, die explizit bayesianische Elemente enthält. Auch diese Konzeption wollen wir im Vergleich zu der Tukeyschen kurz ansprechen. Eine Theorie der explorativen Datenanalyse¹⁷, die diese Vorgehensweise als Synthese aus Deskription und Inferenz auffaßt, ist bislang zwar erst in Ansätzen entwickelt. Wir denken aber, daß die geübte Praxis der Modellkonstruktion aufgrund vorhandener Daten einer theoretischen Begründung bedarf und müssen daher eruieren, inwieweit eine – richtig verstandene – explorative Datenanalyse dazu beitragen kann. Gerade für die Zeitreihenanalyse und hier speziell für eine Modellklasse, die man als ARIMA- oder Box-Jenkins-Modelle bezeichnet, kommt diesem Punkt zentrale Bedeutung zu. Hier ist das theoriefreie Vorgehen bzw. die Modellevaluierung anhand von gegebenen Daten geradezu Programm.

Erst in jüngster Zeit sind eine Reihe von Verfahren entwickelt worden, die dem Aspekt der Modellkonstruktion Rechnung tragen sollen. Hier wollen wir zumindest ausblickhaft auf einige unserer Ansicht nach vielversprechende Aspekte hinweisen, die in der Zukunft stärker beachtet werden sollten.

¹⁷ I. J. Good (1983b) sprach in diesem Zusammenhang sogar von einer “Philosophie” der explorativen Datenanalyse. Dies dürfte wohl zu weit gehen. In Deutschland wird eine wissenschaftstheoretische Sicht der Konzeption vor allem von Schnell (1994) vertreten, allerdings mit der Konsequenz eines Verzichts auf Signifikanztests. Siehe dazu unsere Kritik in Rahlf (1995a, S. 163f).

2. Deskription

a. Allgemeines

Wenn man “die Statistik” als eine Methodenwissenschaft auffaßt, deren Hauptziel ein Erkenntnisgewinn über Massenerscheinungen ist, so steht die *Beschreibung* von Massenerscheinungen sicher an erster Stelle. Ob es sich hierbei lediglich um einen Zwischenschritt oder bereits um das Endziel handelt, kann durchaus als prinzipielle, fachunabhängige Frage gestellt werden. So war etwa G. U. Yule, einer der tiefundesten Statistiker der Moderne, dem wir zahlreiche methodologische Konzepte verdanken, der Ansicht:

“The initial problem of the statistician is simply the description of the data presented; to tell us what the data themselves show. To this initial problem the function of sampling theory is in general entirely secondary or ancillary; to inform the investigator as to the limits within which his descriptive measures can be trusted, so far as fluctuations of simple sampling alone are concerned. The development of theory during my own lifetime followed at first the natural course suggested by this fact. Primarily it was new methods that were developed, and investigations of the ‘probable errors’ involved only followed in their train. More recently methods, with few exceptions (time series in economics, factor-methods in psychology), have been almost neglected, while there has been a completely lopsided development – almost a malignant growth – of sampling theory. I hope there may be a swing back towards the study of method proper, and as methods only develop in connection with practical problems, that means a swing back to more practical work and less pure theory [...]. There are quite large fields of statistics into the discussion of which sampling theory hardly enters at all [...]. Even in the field of experimental work, if the investigator possesses caution, common sense and patience, those qualities are quite likely to keep him more free from error in his conclusions than the man of little caution and common sense who guides himself by a mechanical application of sampling rules. He will be more likely to remember that there are sources of error more important than fluctuations of sampling [...]. No: I cannot assign the place of highest importance to sampling theory – a high place perhaps, but not the highest.”¹⁸

Auch wenn man den Geltungsanspruch der deskriptiven Statistik nicht so universell sehen möchte, so steht doch sicher ihre eigenständige Existenzberechtigung außer Frage. Problematischer gestaltet sich schon die Frage nach den erkenntnistheoretischen Grundlagen oder dem “Wesen” der deskriptiven Statistik. Hier müssen wir allerdings feststellen, daß von Seiten der Fachwissenschaft kaum Anhaltspunkte für eine Beantwortung dieser Frage zu erhalten sind.

Allgemein läßt sich der mittels der Analyse von Massenerscheinungen erstrebte Erkenntnisgewinn durch eine geeignete Charakterisierung dieser Erscheinungen erreichen. Das kann geschehen durch

- Ordnung von Ausprägungen nach bestimmten Kriterien (z. B. Größe) und tabellarischer Auflistung,
- graphische Darstellung der Ausprägungen und

¹⁸ Yule (1942, S. 83f). Hervorhebung von uns.

- schließlich durch Reduktion des Datenmaterials in zusammenfassende Kennzahlen. Dabei kann es sich wiederum um Graphiken oder Tabellen handeln.
- Diese Datenreduktion hat nicht zum Ziel, Erkenntnisse über einen unbekannten, festen Parameter zu erlangen, sondern eine Zusammenfassung zu liefern, die plausibel, anschaulich und einfach ist.¹⁹

Im Rahmen der Wirtschaftsgeschichte können wir z. B. zwei Gruppen einer Gesellschaft hinsichtlich ihres Einkommens vergleichen. Dazu bietet sich der Vergleich einer zusammenfassenden Maßzahl an, etwa einer Maßzahl des mittleren Einkommens. Doch welche Maßzahl ist hierfür zu wählen? Wir verfügen über kein statistisches Kriterium, nach dem das arithmetische Mittel dem Median oder dem Modalwert oder einer anderen Maßzahl²⁰ vorzuziehen wäre.²¹ Ähnlich verhält es sich hinsichtlich der Streuungsmaße. Auch diese haben einen Erkenntniswert an sich, ihre Bedeutung reduziert sich nicht auf eventuelle Schätzeigenschaften. So ist etwa der (selten verwendete) einfache durchschnittliche Abstand

$$(1) \quad d = \frac{1}{N} \sum_{i=1}^n |x_i - \bar{x}|$$

ein viel plausibleres und anschaulicheres Streuungsmaß als die (nahezu ausschließlich verwendete) mittlere quadratische Abweichung

$$(2) \quad s = \sqrt{\frac{1}{N} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Die allgemeine Verwendung letzterer hat ihre Ursache lediglich in dem Umstand, daß sie keine unhandliche Betragsrechnung erfordert und als Schätzer für bestimmte Größen "günstigere" Eigenschaften aufweist. Diese Tatsache ist aber im Rahmen der deskriptiven Statistik irrelevant.²²

b. Anfänge der Statistik als Beschreibung

Zweifellos lassen sich statistische Berechnungen im Sinne von Zusammenzählen verschiedenster Daten bis in die graue Vorzeit zurückverfolgen, doch setzte eine deskriptive Statistik in einem – weitgefaßten – wissenschaftlichen Sinne als "Staatsbeschreibung" erst im 16. Jahrhundert ein.²³

Ein bedeutendes Zentrum der Kameralwissenschaften stellte im 18. Jahrhundert die Universität Halle dar, wo unter anderem Martin Schmeitzel [1679-1747] lehrte: "Er hielt in Jena und später in Halle eine Vorlesung unter dem Titel 'collegium

¹⁹ Vgl. Menges/Skala (1973, S. 309).

²⁰ Z. B. dem arithmetischen Mittel der mittleren 50 %.

²¹ In der Inferenzstatistik ist die Anwendung solcher Kriterien, die – in welchem Sinne auch immer – "beste" Kennzahlen liefern sollen, weit verbreitet. Siehe dazu das nächste Kapitel.

²² Insofern würde sich als deskriptive Maßzahl durchaus auch die durchschnittliche Abweichung vom Median oder der Median der Abweichungen anbieten. Zu weiteren Unterschieden zwischen einzelnen Streuungsmaßen siehe Thome (1989, S. 37-43).

²³ Vgl. Menges (1972, S. 4), allgemein John (1884), Walker (1929), Westergaard (1932).

politico-statisticum' und wurde damit zum Namensgeber der Statistik."²⁴ Dessen Schüler Gottfried Achenwall [1719-1772], der Fach und Begriff popularisierte, markierte den Höhepunkt einer Richtung, die unter der Bezeichnung "Universitätsstatistik" eine Kombination von Geschichte, Geographie und Ökonomie vertrat. Diese Richtung konnte sich in Deutschland bis in das 19. Jahrhundert hinein behaupten, wurde jedoch schließlich von einer anderen abgelöst, die aus der Politischen Arithmetik entstand.

Die Politische Arithmetik ist der eigentliche Beginn der modernen (analytischen) Statistik. Im Gegensatz zur beschreibenden Universitätsstatistik war die Verwendung von Zahlen für sie konstituierend. Allgemein wird ihr Beginn in den Arbeiten von John Graunt [1620-1674] gesehen, der, auf die Geburts- und Sterberegister der Stadt London aufbauend, verschiedene demographische Gesetzmäßigkeiten ableitete. Die Bezeichnung der Richtung stammte von Graunts Zeitgenossen W. Petty [1623-1687], der an die Arbeiten Graunts anknüpfte und sie vorrangig in Hinblick auf ökonomische Phänomene weiterentwickelte.

In England hatte sich bereits im 17. Jahrhundert eine Konsumstatistik entwickelt. 1699 veröffentlichte C. Davenant [1656-1714] einen Essay, in dem er erstmals auf ein Phänomen aufmerksam machte, das man heute als Preiselastizität bezeichnet: Bei einer schlechten Ernte würde der Getreidepreis überproportional steigen. Gut hundert Jahre später wurde diese Regel Gregory King [1648-1712] zugesprochen und gilt seither als "Kingsche Regel".²⁵ In Deutschland begann diese Entwicklung etwa fünfzig Jahre später. Als sich 1752 der Politische Arithmetiker Johann Friedrich Unger [1714-1781] in seiner Schrift *Von der Ordnung der Fruchtpreise und deren Einflüsse in die wichtigsten Angelegenheiten des menschlichen Lebens* mit Getreidepreisschwankungen genauer beschäftigte, setzte er den Grundstein für eine Reihe von weiteren Arbeiten, die sich ebenfalls diesem Thema widmeten.²⁶

Ungers Erkenntnisinteresse ist durchaus typisch für die Politische Arithmetik und zeigt die geistige Verwandtschaft zu seinem Zeitgenossen Johann Peter Süßmilch: Aufgrund von statistischer Beobachtung sollte eine Regelmäßigkeit in der Natur festgestellt werden, die auf die "göttliche Ordnung" zurückzuführen sei und diese wiederum beweise. Das Erkennen dieser Regelmäßigkeiten sollte zur besseren Prognose und damit zum Wohle aller dienen.²⁷

In den ökonomischen Arbeiten des 18. Jahrhunderts finden sich nun auch zunehmend statistische Übersichten und graphische Darstellungen von Zeitreihen: Adam Smiths berühmtes Werk über den *Wealth of Nations* oder eine Darstellung der

²⁴ Vgl. Menges (1972, S. 5).

²⁵ Vgl. dazu Stigler (1954, S. 103f), Abel (1978, S. 23). In Frankreich hatte Sébastien le Prestre de Vauban 1706/7 sein "Projet d'une dixme royale" veröffentlicht, in dem er, ausgehend von einer großangelegten empirischen Untersuchung der sozialen Verhältnisse Frankreichs im 17. Jahrhundert, ein einheitliches Steuersystem und weitreichende soziale Reformen vorschlug. Das Werk liegt jetzt in deutscher Übersetzung vor; siehe Toepel (1994).

²⁶ Das von der Hausväterliteratur ausgehende, vorrangige Interesse an praktischen Arbeitsanleitungen blieb zumindest in Deutschland bis in das 18. Jahrhundert hinein dominierend. Vgl. zu diesem Abschnitt auch Rahlf (1996d, S. 54ff).

²⁷ Vgl. Schmitt (1967, S. 201, 204).

Einfuhren nach und Ausfuhren aus England von 1700 bis 1782 durch W. Playfair aus dem im Jahre 1781 publizierten Werk *The commercial and political atlas*.²⁸

Die historische Entwicklung der Verwendung bestimmter Methoden zur zusammenfassenden Charakterisierung von Zahlenmaterial ist, soweit wir sehen, bislang kaum beachtet worden. Daß auch in diesem uns besonders selbstverständlich erscheinenden Bereich keine einheitliche Entwicklung festzustellen ist, mag das Beispiel des "Mittel"wertes illustrieren. Die Wahl des arithmetischen Durchschnittes aller vorhandenen Daten zur Messung einer unbekannten Größe war keineswegs eine zu allen Zeiten geübte Praxis, wie Eisenhart betont:

"The practice of taking the arithmetic mean of two or more measurements or observed values of a single quantity as the value of the quantity indicated by these measurements or observations seems to have originated in Western Europe more or less simultaneously in several fields – astronomy, metallurgy, navigation – in the latter half of the sixteenth century. Until then the almost universal practice seems to have been to select from among the measurements or observations in hand the one considered to be the 'best' [...]"²⁹

Eine Ausnahme von dieser Regel bildeten moslemische Wissenschaftler im 11. Jahrhundert, die den mittleren Wert zwischen zwei Extremwerten als beste Messung ansahen. Wenn Eisenhart das Aufkommen des arithmetischen Mittels in das 16. Jahrhundert datiert, so ist damit keine Massenerscheinung, sondern lediglich der vereinzelte Gebrauch in wissenschaftlichen Zusammenhängen gemeint. Noch bis in das 18. Jahrhundert hinein war die Verwendung anderer "Statistiken" allgemein verbreitet.³⁰

c. Der Fehlerbegriff und die Entwicklung des Kleinstquadratverfahrens

Das sogenannte Kleinstquadratverfahren gehört heute zweifellos zu den zentralen Verfahren der Statistik. Ziel einer solchen Vorgehensweise ist die Bestimmung einer Größe als (Linear-) Kombination anderer beobachtbarer Größen, um somit eine gegenüber der Realität vereinfachte (Abhängigkeits-) Struktur festzustellen, die zum Beispiel zum besseren Verständnis eines Phänomens beitragen kann oder auch Prognosen einer Größe bei Kenntnis von mit ihr verbundenen Größen erlaubt. Eine Beobachtung y_i kann dann als Linearkombination

$$(3) \quad y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \dots$$

verschiedener anderer Größen $x_{1,i}, x_{2,i}, \dots$ angesehen werden. In der Regel werden jedoch nicht alle Faktoren ermittelt oder fehlerfrei gemessen werden können, die eine Größe y "determinieren". In diesem Fall gilt das Gleichheitszeichen nur, wenn die rechte Seite noch um eine nicht direkt beobachtbare "Rest-" oder "Fehlergröße" ε ergänzt wird.

²⁸ Vgl. die Abbildung in Heiler (1991).

²⁹ Eisenhart (1983, S. 530f). Hervorhebung im Original.

³⁰ Ein Beispiel hierfür liefert die Notierungspraxis auf dem Kölner Getreidemarkt bis in das 18. Jahrhundert hinein. Vgl. Rahlf (1996d, S. 98ff).

In dieser Form findet sich die Gleichung in nahezu allen Lehrbüchern der Statistik, gleich welcher Fachrichtung, als zentrales Analysemittel für die Untersuchung von Einflußfaktoren und Abhängigkeiten bei metrischen Daten. Ebenso findet sich dort das Kleinstquadratverfahren als Methode zur Bestimmung der unbekannten Koeffizienten β_0, β_1, \dots . In der Regel werden den Fehlertermen bestimmte stochastische Eigenschaften unterstellt (zumeist eine Normalverteilung) und das Kleinstquadratverfahren als Teil der Inferenzstatistik aufgefaßt. Gehen wir jedoch zunächst von einem deskriptiven Erkenntnisziel aus. Auch dann lassen sich bereits eine Reihe von Fragen stellen, wie etwa:

- Warum sollten Phänomene über ein starres Gerüst miteinander verbunden sein?
- Warum sollte dies für alle Werte gelten (hohe und niedrige)?
- Warum sollte die Beziehung linear und additiv sein?
- In welchem Sinne sollten die Kleinstquadratkoeffizienten die besten Koeffizienten sein?

Wir wollen im folgenden die Schritte, die zu dieser Entwicklung geführt haben, kurz auflisten.³¹ Hauptsächlich drei Probleme haben sie gekennzeichnet: die Bestimmung der Mondbewegung, die Bewegungen des Jupiter und Saturn und die Ermittlung der genauen Form der Erde.

Ein erster Schritt wurde 1750 von T. Mayer vollzogen. Mayers *Abhandlung über die Umwälzung des Monds um seine Axe und die scheinbare Bewegung der Mondsflecken* gehörte zu den am meisten gelesenen Arbeiten über das Thema. Mayer ging nicht davon aus, daß seine Methode auf andere Probleme übertragbar war oder war zumindest nicht daran interessiert. Das Verfahren wurde jedoch 1771 von J. J. Lalande einer breiten Leserschaft bekannt gemacht, unter anderem auch P. S. Laplace.³² Die Ausgangssituation war, daß man eine Reihe von Gleichungen mit mehreren Unbekannten hatte, die Anzahl der Gleichungen aber die Anzahl der Unbekannten überstieg und keine eindeutigen Lösungen zu bestimmen waren.³³ Sowohl Mayer als auch Laplace fanden zunächst heuristische Lösungen, indem sie die Anzahl der Gleichungen auf unterschiedliche Weise reduzierten. Beide Methoden waren in gewisser Weise ad hoc und kannten kein "best-fit"-Kriterium; dennoch waren sie ein Fortschritt.³⁴

³¹ Wir stützen uns dabei vor allem auf Stigler (1986).

³² Vgl. ebda., S. 31.

³³ Die genaue Messung der Mondbewegung war 1. von großem wissenschaftlichen Interesse, da sich damit die Newtonsche Gravitationstheorie stützen oder widerlegen ließ, 2. von großem kommerziellen und militärischen Nutzen. Eine möglichst genaue Positionsbestimmung auf See konnte sich als unmittelbarer ökonomischer und strategischer Vorteil erweisen.

³⁴ Ein Hauptproblem, das die Durchsetzung eines solchen Verfahrens lange Zeit verhindern konnte, war die unterschiedliche Sichtweise der Statistik und Mathematik. So war es dem Mathematiker Euler bei einem ähnlichen Problem nicht gelungen, eine Lösung zu finden, da er nicht wie ein "Statistiker" annahm, daß sich (zufällige) Fehler gegenseitig ausgleichen, sondern als Mathematiker von einer Fehlerfortpflanzung ausging. Vgl. Stigler (1986, S. 28). Den entscheidenden konzeptionellen Unterschied zwischen dem Ansatz von Mayer und

Die sog. "Mayer-Methode" in ihrer originalen oder in der von Laplace weiterentwickelten Form, die im Zusammenhang mit Problemen hinsichtlich der Umlaufbestimmung der Planeten Jupiter und Saturn sowie des Mondes entstanden waren, fand im Rahmen der Astronomie und Geodäsie allgemeine Verbreitung, auch nachdem das Kleinstquadratverfahren eingeführt worden war. Zwar sah man ihre Nachteile:³⁵ Sie lieferte vor allem keine "besten" Ergebnisse in einem mathematischen Sinne, wenn es darum ging, die beste Kombination aus einer Reihe von inkonsistenten Gleichungen zu finden; statt dessen lieferte sie Ad-hoc-Werte. Man vermutete deren Richtigkeit lediglich intuitiv und vertraute ihr aufgrund erfolgreicher Anwendungen in der Vergangenheit.³⁶ Dennoch waren mit ihr bedeutende Errungenschaften verbunden: Mayers Daten der Mondbewegung sowie seine Karten wurden allgemein als die genauesten angesehen, und schließlich war es Laplace mit eben dieser Methode gelungen, die ins Wanken geratene Newtonsche Gravitationstheorie zu retten und die Bewegungen des Jupiter und Saturn mit ihr in Einklang zu bringen.³⁷ Trotzdem wuchs mit den Jahren auch die Kritik. Zum einen wurden einfachere Wege der Berechnung eines "Ausgleiches" gefunden, zum anderen konnte trotz mehrerer Versuche keine formale Begründung für die Verwendung der "Mayer-Methode" gefunden werden.

Das dritte inhaltliche Problem, das mit der Entwicklung der Ausgleichsrechnung verbunden war, war die Form der Erde.³⁸ Seit dem Ende des 17. Jahrhunderts war aufgrund von Pendelexperimenten bekannt, daß die Erde keine exakte Kugelform aufwies. Newton lieferte hierfür eine Erklärung, die auf der Erdrotation basierte und eine gestauchte Längsachse implizierte. Domenico Cassini, Direktor des Königlichen Observatoriums in Paris, war dagegen der Ansicht, daß die Erde in

Laplace sieht Stigler (1986, S. 37) in der Art der Aggregation. Laplace war der erste, der Beobachtungen *verschiedener Herkunft*, d. h. nicht von einer Beobachtungsreihe, sondern von verschiedenen Beobachtern aus verschiedenen Jahren kombinierte.

³⁵ Vgl. zum folgenden ebda., S. 39.

³⁶ Besonders deutlich wird diese Einstellung in einem Text von Legendre. Noch 1798 erläuterte er für einen Fall, bei dem drei Unbekannte anhand von vier Gleichungen bestimmt werden sollten, daß man nicht die ersten drei lösen und die gefundenen Werte in die vierte Gleichung einsetzen solle, um aus dieser die Fehler zu erhalten, sondern: "Rather we need to try to balance the errors in such a way that they are borne nearly equally by all four equations [...]". Dies könne durch einfaches Ausprobieren geschehen: "[...] this will not be difficult when numerical values have been substituted in the equations." Zitiert nach Stigler (1986, S. 56).

³⁷ Das Problem besteht darin, daß die Umlaufbahn der Planeten durch die Gravitation aller anderen Planeten beeinflusst wird und die Bestimmung der exakten Umlaufbahnen von n Planeten für $n > 2$ einen immensen Rechenaufwand bedeutet. Auch wenn man von der Möglichkeit einer exakten Messung ausgeht, kann die Umlaufbahn praktisch nur approximativ bestimmt werden. Bis heute ist es nicht möglich, die genauen Gründe für Abweichungen vom elliptischen Verlauf nach deterministischen (Vernachlässigung des Einflusses anderer Körper, Sonnendruck etc.) und stochastischen (Meßfehler, atmosphärische Turbulenzen, interstellarer Staub etc.) Faktoren zu trennen. Aufgrund der besonderen Konstellation der Planeten in unserem Sonnensystem ergab sich bei den Messungen eine relativ geringe Fehlervarianz, so daß die *Verteilung* der Fehler keine große Rolle spielte. Vgl. Kalman (1982b, S. 192). Dieser Faktor ist für die methodologische Entwicklung sicher von kaum zu überschätzender Bedeutung.

³⁸ Vgl. zum folgenden Stigler (1986, S. 39ff).

der Längsachse gestreckt sei. Eine Klärung sollten exakte Messungen liefern. Neben Pendelexperimenten konnte dies vor allem durch Winkelmessungen geschehen, bei der eine Linearkombination von drei Werten die Antwort lieferte, und mehrere Messungen gaben schließlich Newton Recht. Die Frage, die noch offen blieb, war das genaue Ausmaß der Abflachung, das aus den nicht exakt gemessenen Winkeln ermittelt werden mußte. Nach mehreren Anläufen lieferte 1760 R. J. Boscovich eine Bestimmungsmethode, die nicht wie diejenige Mayers auf einem Ad-hoc-Vorgehen basierte, sondern von Forderungen ausging, die ein auf den Messungen basierender mittlerer Wert haben sollte. Vor allem sollten sich positive und negative Abweichungen ("Korrekturen") die Waage halten (d. h. die Summe aller Abweichungen sollte Null sein), und die Summe aller absoluten Abweichungen sollte so klein wie möglich sein. Boscovich lieferte ein Berechnungsverfahren und wandte dies auf fünf Messungen an, um die drei gesuchten Unbekannten zu erhalten, doch die Methode fand weder Weiterentwicklungen noch weitere Anwendungen.

Mehrere Jahre später widmete sich Laplace noch einmal dem Problem. Zunächst lieferte er eine analytische Lösung für Boscovichs Ansatz (dieser war geometrisch). Anschließend verwendete er vier der fünf Beobachtungen Boscovichs und schlug vor, den maximalen Fehler zu minimieren. Dieses Vorgehen erwies sich jedoch als zu rechenaufwendig, so daß er eine andere Lösung vorschlug. Die Boscovich-Methode könne verbessert werden, indem die einzelnen Beobachtungen unterschiedlich gewichtet würden.

Es fanden sich im folgenden zwar verschiedene (astronomische und geophysikalische) Anwendungen dieser Methode, doch blieb der große Erfolg aus.

Ein Durchbruch gelang schließlich 1805 A. M. Legendre [1752-1833]. Legendre formulierte in einem Anhang seiner Schrift eine "Methode der kleinsten Quadrate".³⁹

"Of all the principles that can be proposed for this purpose, I think there is no more general, more exact, or easier to apply, than that which we have used in this work; it consists of making the sum of the squares of the errors a minimum. By this method, a kind of equilibrium is established among the errors which, since it prevents the extremes from dominating, is appropriate for revealing the state of the system which most nearly approaches the truth."⁴⁰

Die Anwendung dieses Prinzips führte zu einem "Normalgleichungssystem", das ebenso viele Unbekannte wie Gleichungen enthielt und mit den üblichen mathematischen Mitteln gelöst werden konnte. Zur Illustration lieferte er wieder ein Anwendungsbeispiel, das drei Unbekannte in fünf Gleichungen enthielt.

In den folgenden Jahren breitete sich das Verfahren rasch aus und wurde zur Standardmethode in der Astronomie und Geodäsie.⁴¹ Unmittelbar darauf erfolgte

³⁹ Adrien Marie Legendre, *Nouvelles méthodes pour la détermination des orbites des comètes*, Paris 1805. Die Methode wurde im Zusammenhang mit einem Problem der Erdvermessung eingeführt.

⁴⁰ Zitiert nach Stigler (1986, S. 13).

⁴¹ "The rapid geographic diffusion of the method and its quick acceptance in these two fields, almost to the exclusion of other methods, is a success story that has few parallels in the history

eine systematische Einbettung in einen wahrscheinlichkeitstheoretischen Rahmen. Bevor es jedoch dazu kommen konnte, mußte die Wahrscheinlichkeitstheorie einen Stand erreichen, der diese Übertragung ermöglichte. Diese Übertragung war dann mit einer weiteren fundamentalen konzeptionellen Entwicklung verbunden: der Inversion des Wahrscheinlichkeitsbegriffes. Der Weg dorthin ging zunächst über Anwendungsprobleme der Binomialverteilung. Bevor wir uns dieser Entwicklung zuwenden jedoch einige grundsätzliche Bemerkungen zu diesem Komplex.

3. Inferenz

a. Allgemeines

Die statistische Inferenz ist wie kaum ein anderes Gebiet der Wissenschaft ambivalent. Auf der einen Seite, und diesen Eindruck verstärken auch Blicke in die einschlägigen Lehrbücher (fachübergreifend), ist das heutige Lehrgebäude wie in kaum einem anderen Gebiet über alle Fächergrenzen hinweg konsistent und stringent nach dem gleichen Schema aufgebaut.⁴² Es wird der Eindruck vermittelt, daß es sich um ein logisch strukturiertes, in sich geschlossenes Gebäude handelt, das universale Gültigkeit besitzt. Zwar kann die Gewichtung einzelner Verfahren von Fach zu Fach verschieden sein,⁴³ ihre inferenzstatistischen Grundlagen (d. h. vor allem die Art der Überprüfung von Hypothesen und ganz allgemein die Beurteilung der "Evidenz" statistischer Daten) besitzen aber scheinbar universale Gültigkeit.

Die bis heute diskutierten Inferenzkonzeptionen und damit verbundene Aspekte lassen sich durch die folgenden Themenkomplexe charakterisieren:

- die von K. Pearson ausgebaute Signifikanztestkonzeption,
- die von R. A. Fisher geforderten Kriterien "guter" Schätzungen,
- eine quantitative Bewertung der Güte dieser Schätzungen durch das ebenfalls von Fisher entwickelte Fiduzialprinzip,

of scientific method." Stigler (1986, S. 15). Dennoch konnten sich andere Ansätze, wie etwa derjenige Boscovichs, mit dem Argument halten, der Kleinstquadrat-Ansatz gewichte "Ausreißer" zu stark. Vgl. ebda., S. 55.

⁴² Hier trifft Kuhns Kritik sicher in ähnlicher Weise zu, wie in der von ihm thematisierten Physik. Vgl. Kuhn (1976, S. 25).

⁴³ In der Psychologie haben etwa varianzanalytische Verfahren größere Bedeutung, als in den Wirtschaftswissenschaften oder der Soziologie. Dafür werden im Rahmen der Makroökonomie vor allem regressions- und zeitreihenanalytische Verfahren verwendet, in der Soziologie intensiver loglineare oder Pfadmodelle.

- das – wiederum von Fisher entwickelte und von G. Barnard und A. Birnbaum ausgebaut – Likelihood-Prinzip,
- die “klassische” Testtheorie und Konfidenzinferenz nach J. Neyman und E. S. Pearson,
- das Konzept der strukturellen Inferenz von D. A. S. Fraser,
- subjektivistische Bayes-Ansätze, wie etwa von F. P. Ramsey, B. de Finetti, H. Jeffreys oder L. J. Savage und
- objektivistische Bayes-Ansätze, wie etwa von H. Robbins.

Ein Blick “in das Innere” der Statistik bietet ein anderes Bild. Einige Zitate aus der statistischen Literatur mögen die Kontroversen innerhalb der Fachwissenschaft illustrieren. So konstatierte etwa R. A. Fisher (1956, S. 9), daß “[...] die Theorie der inversen Wahrscheinlichkeit auf einem Irrtume beruht und vollständig abgelehnt werden muß.” Von Mises (1951, S. 188) bekannte bezüglich Fishers Likelihood-Ansatz: “Die vielen schönen Worte, die Fisher und seine Anhänger zur Begründung der likelihood-Theorie beibringen, sind mir nicht verständlich. Das Hauptargument [...] sagt mir nichts.” A. Birnbaum, der das Likelihood-Konzept 1962 in einem vielbeachteten Beitrag zu einem Likelihood-Prinzip als fundamentale Grundlage der Inferenzstatistik erhob, lehnte das von J. Neyman entwickelte Konfidenzprinzip ab, weil es gegen das Likelihood-Prinzip verstoße. Wenige Jahre später ging er jedoch dazu über, das Likelihood-Prinzip abzulehnen, weil es gegen das Konfidenzprinzip verstoße.⁴⁴ Stegmüller (1973, S. 2) verweist auf Neyman, der behauptet habe, daß die von Fisher entwickelten Testmethoden “[...] in einem mathematisch präzisierbaren Sinn ‘schlechter als nutzlos’ seien [...]”⁴⁵ B. de Finetti, einer der Hauptvertreter einer subjektivistischen Wahrscheinlichkeitstheorie, war überzeugt, daß Fisher “[...] sein Gefühl für die Notwendigkeit der Konklusion in Bayes’scher Form zeigte (mit der Illusion, sie mit einer undefinierbaren ‘fiducial probability’ ausdrücken zu können) und das Problem in einer Bayes entgegengesetzten Form darstellen wollte (im wesentlichen wie Neyman).” L. J. Savage, ein weiterer bedeutender Vertreter einer subjektivistischen Sichtweise, der in seinem einflußreichen Buch *The Foundations of Statistics* die konventionellen inferenzstatistischen Verfahren in ein von ihm entwickeltes Axiomensystem einer subjektivistischen Lehre eingliedern wollte, schrieb in der zweiten Auflage seines Buches: “Freud alone could explain how the rash and unfulfilled promise (made early in the first edition, to show how frequentist ideas can be justified by means of personalistic probabilities) went unamended through so many revisions of the manuscript.”⁴⁶ O. Kempthorne (1971) schließlich charakterisierte die verschiedenen Inferenzkonzeptionen auf eine Weise, die J. W. Pratt (1971, S. 496) in einem Kommentar zu der folgenden Zusammenfassung der Thesen Kempthornes veran-

⁴⁴ Vgl. Birnbaum (1962), Birnbaum (1968) sowie Birnbaum (1977).

⁴⁵ Hervorhebungen im Original.

⁴⁶ Zitiert nach DuMouchel (1992, S. 527). Die erste Auflage erschien 1954. Vgl. Savage (1954).

laßte: “Fiducial and structural methods are nonsense. Jeffrey’s Bayesian and subjective Bayesian methods are nonsense. Likelihood methods are nonsense. He doesn’t say directly that orthodox methods are nonsense but he says it implicitly by his remarks [...]. In short, he says all methods are nonsense, therefore use orthodox methods.” Diese Liste ließe sich ohne weiteres verlängern, doch mögen die Eindrücke genügen.⁴⁷

Gerade auf diesem Gebiet, auf dem die äußere mathematische Form auch eine innere Logik suggeriert, beweist sich unseres Erachtens die besondere Tragfähigkeit einer wissenschaftshistorischen Vorgehensweise. Die konzeptionelle Entwicklung von der Wahrscheinlichkeit zur Inferenz, d. h. zu einer Quantifizierung der Unsicherheit eines Beobachtungsfundes ging zwei Wege: erstens über die Binomialverteilung und zweitens über den Fehlerbegriff. Diesen Wegen und den darauf aufbauenden Konzeptionen wollen wir im folgenden Abschnitt nachgehen.

b. Entstehung der Inferenzstatistik

Wir wollen zunächst die Entwicklung zentraler Konzepte in groben Zügen darstellen. Folgende Begriffe markieren die wichtigsten Schritte:

<i>Zeit (ca.)</i>	<i>konzeptionelle Entwicklung</i>
1700 – 1730	die ersten systematischen Definitionen der Begriffe “Wahrscheinlichkeit” und “Zufall” (Leibniz, Bernoulli); der Versuch, von der Wahrscheinlichkeitsrechnung zur Inferenzstatistik (Rückschluß) zu gelangen (Bernoulli);
1750 – 1775	die Inversion des Wahrscheinlichkeitsbegriffes im Zusammenhang mit der Fehlerrechnung (nach Vorarbeiten von Cotes, Simpson, Lagrange, Lambert, D. Bernoulli durch Laplace); die Inversion des Wahrscheinlichkeitsbegriffes im Zusammenhang mit der Binomialverteilung durch Bayes;
um 1810	die Synthese aus Fehler- und Wahrscheinlichkeitsrechnung durch Gauss und Laplace;
1820 – 1840	die Übernahme und Weiterentwicklung inferenzstatistischer Konzepte (Fehlergesetz, Gesetz der großen Zahlen) in die “Sozialwissenschaft” durch Quetelet;
1870 – 1885	die Übernahme der Konzepte Quetelets in die Biologie durch Galton und konzeptionelle Grundlegung der Korrelation und Regression;

⁴⁷ In der gängigen Lehrbuchliteratur finden sich in der Regel keine Hinweise auf die Existenz dieser Kontroversen. Die statistische Spezialliteratur ist dagegen zahlreich und kaum systematisch zu überblicken, da einzelne Positionen oftmals nicht klar voneinander abgegrenzt sind. Einen ersten Eindruck von der Breite und Tiefe des Spektrums vermitteln etwa die Diskussionen von Savage et al. (1962) oder die Sammelbände von Godambe/Sprott (1971) und Harper/Hooker (1976). Kurze, aber prägnante Überblicke gibt Menges (1967a) und Menges (1972, S. 265-286). Aus wissenschaftstheoretischer Sicht haben vor allem Hacking (1965) und Stegmüller (1973) versucht, sich dem Thema zu nähern.

1880 – 1895	die Systematisierung und Formalisierung inferenzstatistischer Konzepte durch Edgeworth und K. Pearson;
1895 – 1900	die Anwendung dieser systematisierten inferenzstatistischen Konzepte auf sozialwissenschaftliche Daten und Weiterentwicklung zur multiplen Regression durch Yule; schließlich
1895 – 1900	der Versuch einer Klärung der Begriffe Korrelation, Scheinkorrelation und Kausalität durch K. Pearson und Yule.

Inferenzaussagen sind zunächst Aussagen über Wahrscheinlichkeiten.⁴⁸ Deren systematische Untersuchung läßt sich zweifelsfrei bis in das 13. Jahrhundert zurückverfolgen und war motiviert durch eine Berechnung von Wahrscheinlichkeiten für Ereignisse in Glücksspielsituationen. Die Wahrscheinlichkeitstheorie entstand aus dem Versuch, "Probleme aus dem Bereich der Glücksspiele mit mathematischen Mitteln zu lösen".⁴⁹ Dabei ergaben sich keine Probleme mit einer Definition des Begriffes 'Wahrscheinlichkeit', denn hier hatte man es immer mit der einfachen Situation von abzählbar vielen Ereignissen mit jeweils gleich "wahrscheinlichen" Möglichkeiten zu tun. Rein intuitiv lag es nahe, jeder Augenzahl bei einem Würfelwurf oder jeder Seite beim Wurf einer Münze die gleiche Wahrscheinlichkeit zuzuschreiben.

Die erste praktische Anwendung des Wahrscheinlichkeitsbegriffes außerhalb von Glücksspielsituationen fand in bezug auf soziale Phänomene, insbesondere auf Sterblichkeitsprobleme statt, in denen man ebenfalls Gesetzmäßigkeiten zu entdecken glaubte. Konsequenterweise erfolgte hier eine Gleichsetzung von Wahrscheinlichkeiten und Wettquotienten. Eine der ersten Arbeiten dieser Art ist die 1662 von John Graunt [1620-1674] veröffentlichte Untersuchung über Gesetzmäßigkeiten in der Bevölkerungsentwicklung anhand der Totenregister von London. In einem Brief von Christiaan Huygens an seinen Bruder Lodewijk, in dem die Arbeit Graunts aufgegriffen wurde, tritt die zu dieser Zeit vorherrschende Auffassung beispielhaft zutage:

⁴⁸ Wir beschränken uns hier auf die Nennung der wichtigsten Literatur. Die historische Entwicklung zentraler Konzepte der Statistik ist in einer Reihe von Beiträgen in der Zeitschrift *Biometrika* untersucht worden, die in Pearson/Kendall (1970) wiederabgedruckt wurden. Diesem Band folgte Kendall/Plackett (1977), der eine Reihe weiterer verstreuter Aufsätze zusammenfaßt. Die Diffusion wahrscheinlichkeitstheoretischer Konzepte in die einzelnen Wissenschaften wird in den Sammelbänden von Krüger/Daston/Heidelberger (1987) und Krüger/Gigerenzer/Morgan (1987) behandelt. Zahlreiche Beiträge zur Entwicklung der Statistik hat in den letzten Jahren S. M. Stigler geliefert, die er 1986 in einer bahnbrechenden Monographie (Stigler (1986)) zu einer Synthese zusammenführte. Dort finden sich auch kommentierte weiterführende Literaturhinweise. Den Transfer theoretischer Konzepte zwischen den Wissenschaften von 1820 bis 1900 behandelt Porter (1986). Eine enzyklopädische Darstellung der historischen Entwicklung inverser Inferenztechniken und -konzepte von Thomas Bayes bis Karl Pearson verdanken wir Dale (1991). Für die historische Entwicklung der Wahrscheinlichkeitstheorie bis 1933 (Formulierung der Kolmogoroff-Axiome) bietet die Quellensammlung von Schneider (1988) eine unverzichtbare Grundlage, für bahnbrechende Arbeiten der theoretischen Statistik des 20. Jahrhunderts diejenige von Kotz/Johnson (1992).

⁴⁹ Schneider (1988, S. 1).

“Hier noch ein weiteres Beispiel. Nehmen Sie an, daß ich in Bezug auf 100 Kinder zum Zeitpunkt der Zeugung, wie üblich vorausgesetzt, für jedes von ihnen gewettet hätte, daß es ein Alter von 16 Jahren erreicht. Da gewöhnlich nicht mehr als 40 von 100 16jährigen übrig bleiben, hätte ich sicherlich einen Nachteil und, um die Partie ausgeglichen zu gestalten, dürfte ich nur 40 gegen 60 oder 2 gegen 3 gesetzt haben.”⁵⁰

Nach theologischen Diskussionen im Zusammenhang mit dem Probabilismusstreit zwischen Jansenisten und Jesuiten⁵¹ erfolgte dann seit dem Ende des 17. Jahrhunderts die Einbettung des Begriffes in ein naturwissenschaftliches Weltbild. Zahlreiche Einzelprobleme waren zu diesem Zeitpunkt behandelt worden, doch beschränkten sich diese auf kombinatorische Ableitungen von a priori vorgegebenen Wahrscheinlichkeiten. Die Wahrscheinlichkeitstheorie wäre ein Gedankenspiel ohne große Folgen geblieben, hätte man nicht erkannt, daß sich die bisherigen Überlegungen auch umkehren ließen. Es ließ sich nicht nur nach der Wahrscheinlichkeit für das Eintreten eines Ereignisses oder einer Ereigniskombination fragen, wenn die Wahrscheinlichkeiten a priori bekannt waren, wie etwa beim Münz- oder Würfelwurf, sondern auch umgekehrt: “How, from the outcome of a game (or several outcomes of the same game), could one learn about the properties of the game and how could one quantify the uncertainty of our inferred knowledge of these properties?”⁵² Die Beantwortung dieser Frage, der Rückschluß, bildet die eigentliche Grundlage der *Inferenzstatistik*, die eine quantitative Angabe über den Grad der mit ihm verbundenen Unsicherheit ermöglicht.

Vermutlich der erste, der sich mit dieser Problematik in mathematischer Form auseinandersetzte, war Jakob Bernoulli [1655-1705]. Bernoulli, der mit seiner Arbeit *Ars conjectandi* einen Grundstein der modernen Statistik legte, ging davon aus, daß es für empirische Phänomene feste (konstante) Wahrscheinlichkeiten gäbe, die durch Versuche in theoretisch beliebig genauer Annäherung ermittelt werden könnten – wobei die Annäherung um so besser sei, je mehr Fälle man betrachte. Zwar hatte es auch schon vorher intuitive Vermutungen gegeben, daß mit zunehmender Zahl von Beobachtungen eine unbekannte Proportion immer genauer bestimmt werden kann, doch formulierte Bernoulli erstmals eine formale Behandlung dieser vagen Vermutung, indem er eine Approximationsformel entwickelte, die numerische Grenzen angab, innerhalb derer – mit vorgegebener Wahrscheinlichkeit – sich eine relative Häufigkeit befinden wird.⁵³ Damit war der Grundstein für eine moderne Schätztheorie gelegt: “The conflicting philosophical

⁵⁰ C. Huygens in einem Brief an L. Huygens vom 21.11.1669, zitiert nach Schneider (1988, S. 186).

⁵¹ Interessanterweise spielen laut Schneider (1988, S. 47) neben theologischen auch ökonomische Motive eine Rolle für den Versuch einer Quantifizierung des Wahrscheinlichkeitsbegriffes: “[...] ebenso wichtig dafür war die im Gefolge der Renaissance und der Entwicklung des Handels im 16. Jahrhundert beobachtbare Zukunftsorientierung, die durch den frühkapitalistischen Markt mit seinen Methoden der Spekulation und Manipulation ein wesentlich höheres Gewicht erhielt.”

⁵² Stigler (1986, S. 63).

⁵³ Eine Approximationsformel war von besonderem Interesse, da der Rechenaufwand der Wahrscheinlichkeiten für größere Beobachtungszahlen so aufwendig wurde, daß eine exakte Berechnung praktisch ausgeschlossen war. Vgl. Porter (1986, S. 93).

interpretations of Bernoulli's quantitative statement would come much later, but he had taken the first step on what was to be a long and eventful journey."⁵⁴

Bernoulli war vor allem an der Frage interessiert, wie groß die Anzahl der Beobachtungen sein müßte, damit – bei bekannter wahrer Proportion – die empirische relative Häufigkeit mit vorgegebener Wahrscheinlichkeit ("moral certainty") in ein vorgegebenes Intervall falle. Bezeichnet man die Anzahl der Fälle, in denen ein Ereignis eingetreten ist, mit r (in Bernoullis Diktion als "fruchtbar"), die Anzahl der Fälle, in denen ein Ereignis nicht eingetreten ist, mit s ("unfruchtbar") und die Gesamtzahl der Fälle mit $t = r + s$, so lautet sein Hauptsatz:

"Es können so viele Versuche angestellt werden, daß es sich als beliebig vorgegeben, z. B. c -mal wahrscheinlicher herausstellt, daß die Anzahl der fruchtbaren Beobachtungen innerhalb dieser Grenzen als außerhalb fallen wird, d. h., daß das Verhältnis der Anzahl der fruchtbaren zur Anzahl aller Beobachtungen nicht größer als $(r+1)/t$ und nicht kleiner als $(r-1)/t$ sein wird [...]"⁵⁵

Bernoullis Überlegungen wurden von Abraham De Moivre [1667-1754] weitergeführt, der im Gegensatz zu Bernoulli nicht an der Bestimmung des notwendigen Beobachtungsumfangs N bei vorgegebener Wahrscheinlichkeit und Sicherheit, sondern an der Bestimmung der Wahrscheinlichkeit bei vorgegebenem N interessiert war.⁵⁶ De Moivre lieferte eine auf der Exponentialverteilung basierende Approximationsformel und zeigte, daß die Hauptmasse der Wahrscheinlichkeit in ein Intervall $[p-\epsilon; p+\epsilon]$ fällt, wenn für ϵ ein Wert proportional zu $1/\sqrt{n}$ bei sehr großem n gewählt wird.⁵⁷ Neben einer verbesserten Approximationsformel geht vor allem die quantitative Angabe der Beziehung zwischen Umfang der Beobachtungen und Genauigkeit des Ergebnisses über Bernoulli hinaus: Die Genauigkeit wächst mit der *Quadratwurzel* des Beobachtungsumfangs. Zwar war diese Erkenntnis zunächst nur auf binomiale Situationen beschränkt, doch war ein Anfang gemacht, der später vor allem von Laplace ausgebaut wurde.⁵⁸

Ebensowenig wie Bernoullis Ausführungen fanden De Moivres Ergebnisse unmittelbare Anwendung. Stigler (1986, S. 85f) vermutet folgenden Grund:

"One primary reason, I believe, is that it did not answer the question that is most often and most naturally asked in applications. It did not tell empirical scientists what, given an available empirical fact, they could say about the process that produced the fact. It did not provide an answer to the fundamental question of statistical inference. For example, if De Moivre or someone else of his era knew that of 346 men of age fifty, only 142 survived to age seventy,"⁵⁹ he might ask, 'How much credence can I give to the ratio 142/346 as an estimate of the chance

⁵⁴ Stigler (1986, S. 70).

⁵⁵ Jakob Bernoulli, *Ars conjectandi*, Basel 1713, S. 236 u. 239, zitiert nach Schneider (1988, S. 124). Auf die Wiedergabe der Approximationsformeln wollen wir hier verzichten.

⁵⁶ Abraham De Moivre, *The doctrine of chances*, 2. A., London 1738 [3. Aufl. 1756].

⁵⁷ Vgl. Schneider (1988, S. 118).

⁵⁸ Vgl. Stigler (1986, S. 83f). T. Mayer war im Zusammenhang mit der Fehlerrechnung 1750 noch davon ausgegangen, daß die Genauigkeit mit wachsender Zahl von Messungen proportional steigt.

⁵⁹ Dies ist ein Beispiel, das De Moivre 1725 in der ersten Auflage seines Buches *Annuities on lives* von Halley übernommen hatte.

of surviving from fifty to seventy? What for example, is the probability this number is below 1/2? De Moivre's results did not permit a direct answer to these questions."

Statt dessen beantworteten sie die folgende Frage: *Wenn die wahre Wahrscheinlichkeit 1/2 beträgt, wie groß ist dann die Wahrscheinlichkeit, daß sich ein Wert von 142/346 oder kleiner realisiert?*: "His calculations did not contain even a germ of inverse probability [...]. For De Moivre chance lay in the data, not in the underlying probabilities. The successful mathematical treatment of the inference problem was to require the abandonment of this restrictive view."⁶⁰

Bevor wir uns der Weiterentwicklung des Inferenzproblems zuwenden, müssen wir noch auf ein weiteres Problem hinweisen, das sich in diesem Zusammenhang stellte und auch bis in unsere Zeit Anlaß zu Kontroversen gibt: die Beziehung von Zufall und Gesetz und – damit verbunden – das "Wesen" von Gesetzmäßigkeiten. Bernoullis Vorstellungen fanden in dieser Hinsicht keine ungeteilte Zustimmung. So erwiderte etwa G. W. Leibniz [1646-1716]:

"Das Hauptproblem scheint mir darin zu bestehen, daß zufällige Ereignisse bzw. das, *was von unendlich vielen Umständen abhängt*, nicht durch endlich viele Versuche bestimmt werden kann. Zwar hat die Natur ihre Gewohnheiten, die aus der Wiederkehr der Ursachen erwachsen, aber nur im Regelfall. Wer sagt deshalb, ob nicht der nächste Versuch gerade wegen der Veränderlichkeit der Dinge beträchtlich von der Regel aller vorhergehenden abweicht? Es treten immer wieder neue Krankheiten bei den Menschen auf; wenn man also auch beliebig viele Untersuchungen über die Sterblichkeit gemacht hat, hat man die Grenzen für die natürlichen Dinge keineswegs so festgelegt, daß sich in Zukunft nichts ändern könnte."⁶¹

Zwei Aspekte sind hier von Interesse. Leibniz bezeugte zum einen eine indeterministische Weltsicht. Für ihn war zwar auch ein Ursache-Wirkung-Schema grundlegend, darin war aber die Möglichkeit eingeschlossen, daß es "unendlich viele" Ursachen gäbe, somit echte Zufälligkeit in einem ontologischen Sinne existiere. Zum anderen bestritt er zwar nicht, daß die Natur Gesetzmäßigkeiten unterliege, diese stünden aber gerade im Gegensatz zu zufälligen Ereignissen. Die Erkenntnisprobleme in bezug auf zufällige Ereignisse, die Gesetzmäßigkeiten unterliegen, kommen hier in geradezu idealtypischer Weise zum Ausdruck. Das zweite Problem, das Leibniz anspricht, würde man in moderner Sprache etwa mit "Konstanz von Parametern" bezeichnen. Bezogen auf soziale Phänomene ist diese Konstanz seiner Ansicht nach nicht gegeben. J. Bernoulli nahm zu dieser Kritik in seiner 1713 erschienenen Schrift wie folgt Stellung:

"Zufällig sowohl im Sinn von frei, d. h. von der Willensentscheidung eines rationalen Geschöpfes abhängig, als auch im Sinne des aufs Geratewohl, d. h. des von Nebenumständen bestimmten, ist das, was nicht sein, nicht werden oder nicht gewesen sein könnte, wohlgemerkt aufgrund einer fernliegenden, nicht der nächstliegenden Verwirklichungsmöglichkeit; das Zufällige schließt nämlich nicht immer jede Notwendigkeit hinsichtlich der sekundären Ursachen aus, was ich an Beispielen erklären werde. Ein Würfel kann mit absoluter Sicherheit zu dem Zeitpunkt, zu dem er die Hand des Werfenden verläßt, bei gegebener Lage, Geschwindigkeit

⁶⁰ Stigler (1986, S. 86).

⁶¹ Aus einem Brief von Leibniz an Jakob Bernoulli vom 3.12.1703, in: Schneider (1988, S. 59). Hervorhebung von uns.

und Abstand vom Spielbrett nicht anders fallen, als er tatsächlich fällt. Ebenso wird, wenn der heutige Zustand der Luft, die Masse, Lage, Richtung und Geschwindigkeit der Winde, Dämpfe und Wolken sowie die Gesetze des mechanischen Zusammenhanges gegeben sind, nach denen dies alles miteinander in Wechselwirkung tritt, das Wetter von morgen nicht anders werden können als es wirklich sein wird; diese Wirkungen folgen also aus ihren unmittelbaren Ursachen nicht weniger notwendig als die Verfinsterungserscheinungen aus dem Lauf der Gestirne. Dennoch hat man daran festgehalten, daß man nur Verfinsterungen zu dem Bereich des sich notwendig ereignenden, das Auftreffen des Würfels oder das künftige Wetter aber zu dem Bereich des sich zufällig Ereignenden zählt. *Der Grund dafür ist gerade, daß das, was man zur Bestimmung späterer Wirkungen als gegeben ansieht und was auch in der Natur gegeben ist, uns dennoch nicht hinreichend bekannt ist.* Wenn dies der Fall wäre, wäre die mathematische und physikalische Forschung genügend weit entwickelt, um diese Wirkungen aus den gegebenen Daten berechnen zu können, genauso wie man aus den klar erkannten Grundgesetzen der Astronomie Verfinsterungen berechnen und voraussagen kann, die, bevor die Astronomie zu solcher Vollkommenheit gediehen war, deshalb selbst nicht weniger als die beiden anderen zu den zukünftigen zufälligen Ereignissen gerechnet werden mußten. Daraus folgt, daß einer Person zu einem Zeitpunkt etwas zufällig erscheinen kann, was für eine andere Person – ja sogar für dieselbe – zu einem anderen Zeitpunkt, nach dem Bekanntwerden seiner Ursachen notwendig eintritt. *Zufälligkeit betrifft also auch vor allem unseren Kenntnisstand*, nämlich inwieweit wir im Objekt keinen Widerspruch gegen das gegenwärtige oder zukünftige Nichtsein sehen, obwohl es hier und jetzt kraft der unmittelbaren, uns aber unbekannten Ursache notwendig ist oder wird.“⁶²

Und konkret auf den letzten Punkt der Leibnizschen Kritik bezogen:

“Meine Erwiderung lautet, ich kann nicht leugnen, daß sich im Verlauf der Zeit die Krankheiten vervielfachen können, und sicherlich wird der, der aus heutigen Beobachtungen auf die Zeiten der Patriarchen vor der Sintflut schließen will, von der Wahrheit gewaltig abweichen. Daraus folgt aber nur, daß *gelegentlich neue Beobachtungen angestellt werden müssen*, genauso wie sie bei den Steinchen angestellt werden mußten unter der Annahme, daß sich ihre Anzahl in der Urne ändert.“⁶³

Zufälligkeit war für Bernoulli gleichbedeutend mit einem unzureichenden Kenntnisstand. Ein Ereignis könne nur endlich viele Ursachen haben. Diese endlichen, aber unbekannten Ursachen ließen sich für ihn am besten durch eine Wahrscheinlichkeitsverteilung beschreiben. Das bedeutet, die Ursachenkonstellation ist zwar unbekannt, bleibt aber konstant. Seine direkte Erwiderung an Leibniz zeigte aber auch einen gewissen Pragmatismus, der strenggenommen zu seiner These, Wahrscheinlichkeiten durch Versuche beliebig genau schätzen zu können, in Widerspruch steht. Bernoulli dachte hier eher in mathematischen bzw. theoretischen

⁶² Jakob Bernoulli, *Ars conjectandi*, zitiert nach Schneider (1988, S. 63f). Übersetzung von Schneider, Hervorhebungen von uns.

⁶³ Ebda., S. 68. Hervorhebung von uns. Diese “Zugeständnisse” gab Bernoulli jedoch wieder auf, als er am Ende des Werkes seinen zentralen Satz (s. o.) formulierte. Er kommentierte ihn dort: “Daraus scheint sich diese bemerkenswerte Folgerung zu ergeben, daß wenn die Beobachtungen aller Ereignisse in alle Ewigkeit fortgesetzt würden, wobei schließlich die Wahrscheinlichkeit in vollkommene Sicherheit übergehen würde, alles in der Welt als sich in bestimmten Verhältnissen und nach einer festen Gesetzmäßigkeit des Wandels sich ereignend erkannt würde, so daß wir sogar gehalten wären, selbst bei den zufälligsten Dingen gleichsam eine gewisse Notwendigkeit und sozusagen eine Bestimmung anzuerkennen” (S. 124).

Kategorien. Insofern war der Einwand von Leibniz für ihn auch kein wirkliches Problem. Seiner Meinung nach reichte es aus, wenn "ähnliche Umstände" vorliegen:

"[...] was sich der Bestimmung a priori entzieht, läßt sich schließlich a posteriori, d. h. aus der oftmaligen Beobachtung des Ereignisses unter ähnlichen Umständen ermitteln, weil man annehmen muß, daß jedes <Ereignis> künftig in so vielen Fällen eintreten oder nicht eintreten kann, wie vorher bei einem ähnlichen Stand der Dinge festgestellt wurde, daß es eingetreten oder nicht eingetreten war; denn wenn man z. B. irgendwann im Rahmen einer Untersuchung beobachtet hätte, daß von 300 Menschen desselben Alters und desselben Gesundheitszustandes wie Titius 200 vor Ablauf der nächsten zehn Jahre gestorben sind und die übrigen mehr <als zehn Jahre> weitergelebt haben, könnte man mit ausreichender Sicherheit folgern, daß es doppelt so viele Fälle gibt, in welchen auch Titius innerhalb der nächsten zehn Jahre den Weg alles Irdischen gehen müßte, wie solche, bei denen er dieses Intervall überleben könnte [...]"⁶⁴

Hier deutete sich bereits ein weiteres Problem an, das Wahrscheinlichkeitstheoretiker und Philosophen bis in unser Jahrhundert hinein beschäftigen sollte: Bernoulli sprach hier auf der einen Seite von empirischen relativen Häufigkeiten, die als Schätzer für Wahrscheinlichkeiten dienen (s. o.),⁶⁵ ging aber auf der anderen Seite auch von individuellen Einzelwahrscheinlichkeiten aus. Diesen Schritt wollte er jedoch nicht konsequent zu Ende denken und konstruierte daher fiktive "Fälle", die sich dann wieder als relative Häufigkeit interpretieren ließen.

Die Problematik der Parameterkonstanz in bezug auf soziale Phänomene wurde von De Moivre im Zusammenhang mit Rentenberechnungen noch einmal aufgegriffen. E. Halley [1656-1742] hatte 1693 eine Sterbetafel aufgrund mehrerer Tausend Angaben zu Geburten und Sterbefällen aus Breslau aufgestellt, die wiederum von De Moivre kommentiert wurde. All diesen Untersuchungen war gemein, daß sie von Gesetzmäßigkeiten bezüglich einer Sterbe"wahrscheinlichkeit" (Abraham De Moivre verwandte den Begriff "Lebenswahrscheinlichkeit") ausgingen, die in hinreichender Annäherung durch empirische Daten erfaßt werden könne. Die Abweichungen der Daten von der wahren Größe wurde mit einer nicht ausreichend großen Anzahl von Daten bzw. eines nicht ausreichend langen Beobachtungszeitraums begründet: "Daraus geht hervor, daß die Beobachtungstafel [von E. Halley, T. R.] als das Resultat der Erfahrung nur weniger Jahre, nicht ganz so verläßlich wäre, um sie zur Grundlage einer festen und unabänderlichen Bewertung von Leibrenten zu machen; aber zugegeben, eine solche Tafel wäre erreichbar, indem man sie auf die Erfahrungen einer großen Anzahl von Jahren gründet [...]"⁶⁶

Auf diese Aspekte werden wir später noch einmal zurückkommen, uns zunächst aber kurz der Weiterentwicklung der Quantifizierung der Unsicherheit zuwenden.

⁶⁴ Ebda., S. 65f.

⁶⁵ Diese – aus mathematischer Sicht – unproblematische Konzeption bildete die Grundlage für Bernoullis berühmtes Theorem: das Gesetz der großen Zahlen.

⁶⁶ Abraham De Moivre, *Annuities on lives*, 2. A., London 1743, zitiert nach Schneider (1988, S. 206). De Moivre hat seine Erkenntnisse auf dem Gebiet der Wahrscheinlichkeitsrechnung bezüglich der Abschätzung der Unsicherheit (s. o.) hier nicht verwendet. Vgl. Stigler (1986, S. 85).

Diese Weiterentwicklung vollzog sich jedoch nicht im Rahmen der Binomialverteilungen. Mehrere Gründe verhinderten dies: Erstens war ihre Anwendung nur auf begrenzte Situationen möglich. Zweitens hängt die Verteilung der Abweichung zwischen der empirischen relativen Häufigkeit und dem wahren Verhältnis von eben jenem wahren Verhältnis ab. Daher kann die Verteilung dieser Abweichung nicht als bekannt angenommen werden, solange der wahre Anteilswert (das wahre Verhältnis) nicht bekannt ist. Somit handelt es sich bei Abschätzungen der Unsicherheit immer um Annäherungen, in denen (z. B. für die Varianz) der wahre Anteilswert durch die entsprechende relative Häufigkeit ersetzt werden muß. Und drittens implizierten die Näherungsformeln von Bernoulli und De Moivre so große notwendige Beobachtungszahlen, daß ihre Anwendung auf "interessante" Fragestellungen völlig aussichtslos erschien.⁶⁷ Um die von Bernoulli geforderte Aussagensicherheit ("moral certainty") von $1000/1001=0.999001$ zu erhalten, war nach seinem Beispiel bei einem vorgegebenem Verhältnis⁶⁸ 30/50 für ein Intervall der Breite 1/50 eine Anzahl von 25.550 Beobachtungen nötig. Keine Untersuchung, schon gar nicht im Rahmen der von Bernoulli intendierten Geschlechterverhältnisse, konnte auch nur annähernd diesen Umfang erreichen.

Erst in der Fehlerrechnung waren zumindest die ersten beiden Probleme nicht gegeben. Messungen, die Fehler verursachten, kamen in vielen Situationen vor, und hierbei konnte man von festen vorgegebenen (Fehler-)verteilungen ausgehen: etwa von einem Mittelwert 0 und einer um diesen Mittelwert symmetrischen Verteilung. Vor allem aber waren die konzeptionellen Schwierigkeiten bei der Inversion der Wahrscheinlichkeit im Zusammenhang mit Fehlergrößen viel geringer. Wenn man mit e eine Fehlergröße bezeichnet, mit B eine Beobachtung und mit P den beobachteten Punkt, dann bedeutet $B = P + e$ gleichermaßen, daß $P = B - e$ ist. Wenn man weiter annimmt, daß e eine symmetrisch verteilte Zufallsgröße darstellt, impliziert die Annahme, daß P fest ist, eine Verteilung für B ; nimmt man andererseits B als gegeben, erhält man eine Verteilung für P . Betrachtet man schließlich die Differenz $e = B - P$ als zufällig, entsteht eine symmetrische Situation, in der die

"[...] inversion becomes most natural. R. A. Fisher was to call this a fiducial argument [...]. Statisticians now know that this argument is not so simple as a naive view might have it, that deep and subtle philosophical and mathematical points must be dealt with if this most natural approach is to lead to a coherent theory of inference. But in the mid-eighteenth century the argument was a conceptual liberation. Whether expressed in mathematical symbols or in words, it led naturally to a view in which one distribution – the distribution of errors – provided the random element for both 'forward' (in time) and 'inverse' probability statements. And it suggested an idea of inverse inference, reasoning *probabilistically* from effect (O) [in unserer Notation B , T. R.] to cause (P), that was to bear fruit in other applications as well."⁶⁹

⁶⁷ Vgl. ebda., S. 91, 69f, 77.

⁶⁸ Nach Bernoullis Argumentationsgang muß diese Wahrscheinlichkeit a priori bekannt sein.

⁶⁹ Stigler (1986, S. 101). Hervorhebung im Original.

Dennoch wurde dieser Schritt nicht unmittelbar, sondern erst nach einer Reihe von Vorarbeiten von Pierre Simon Laplace [1749-1827] vollzogen. Wenden wir uns zunächst den Vorarbeiten zu.

Schneider setzt den Beginn der Vorstellung, daß Fehler bei technischen Messungen einem Verteilungsgesetz gehorchen könnten, für die Zeit um 1700 an.⁷⁰ Zu dieser Zeit hatten eine Reihe von Meßinstrumenten eine Qualität erreicht, die ein Bewußtsein für die Genauigkeit von Messungen sowie eine Abschätzung des Fehlers schaffen konnten. Nach Vorarbeiten von R. Cotes [1682-1716] und T. Simpson [1710-1761] untersuchten J. L. Lagrange [1736-1813], D. Bernoulli [1700-1782] und P. S. Laplace eine Reihe weiterer hypothetischer Fehlerverteilungen, die nach Schneider (1988) jedoch rein theoretisch orientiert waren.⁷¹ Erst J. H. Lambert [1728-1777] unternahm 1760 seiner Ansicht nach den Versuch, "die verschiedenen, nur teilweise ein Fehlerbewußtsein berücksichtigenden Methoden in den messenden bzw. beobachtenden Naturwissenschaften mit einer wahrscheinlichkeitstheoretisch gestützten Fehlertheorie zu verbinden".⁷² Stigler sieht dagegen den entscheidenden Durchbruch bereits bei T. Simpson.⁷³

T. Simpson präsentierte der Royal Society 1755 einen Vortrag "*On the advantage of taking the mean of a number of observations in practical astronomy.*" Der entscheidende konzeptionelle Durchbruch dieser Arbeit war nach Stigler, daß Simpson erstmals nicht die Beobachtungen selbst, sondern die Fehler in den Mittelpunkt des Interesses stellte; unter anderem fragte er nach dem mittleren Fehler, nicht nach dem mittleren Beobachtungswert.⁷⁴ Das bedeutete: Auch wenn die genaue Position eines Körpers als unbekannt vorausgesetzt wurde, war die Verteilung der Fehler davon nicht betroffen. Simpson ging nicht von konkreten, sondern von hypothetischen Werten aus, denen er eine Wahrscheinlichkeitsverteilung unterstellte. Sein Ziel war der mathematische Nachweis, daß das arithmetische Mittel die beste Wahl für die Ermittlung der wahren Größe unter der Annahme verschiedener Fehlerverteilungen sei.⁷⁵ Die Fehler selbst postulierte Simpson in einer Form, die eine einfache mathematische Handhabung erlaubten: zunächst in

⁷⁰ Schneider (1988, S. 215): "Die Wurzeln für die Entstehung des Fehlerbegriffs in den exakten Naturwissenschaften und in der Mathematik sind bis heute nicht vollständig freigelegt." Vgl. zum folgenden auch Eisenhart (1983) und Mellor (1982).

⁷¹ Ein Überblick über verschiedene Verteilungen mit Angabe ihrer Urheber findet sich in Eisenhart (1983, S. 532f).

⁷² Schneider (1988, S. 217).

⁷³ "Simpson had seen that the concept of error distributions permitted a back-door access to the measurement of uncertainty. Later Laplace was to slip in this same back door and come around to open the front (only to find later that Bayes's key was already in the lock)." Stigler (1986, S. 94).

⁷⁴ Der Ausgang eines Würfel- oder Münzwurfes ist ungewiß und kann als vom Zufall bestimmt angesehen werden. Eine Messung, etwa die Messung der Länge eines Gegenstandes, kann man dagegen nicht unmittelbar als Versuch mit mehreren möglichen Ausgängen betrachten. Somit ist eine inferentielle Aussage, die sich auf die Länge des Gegenstandes bezieht, nicht möglich. Die Länge des Gegenstandes ist nicht mehr das Ergebnis, sondern die Voraussetzung der Messung. Die Messung kann dagegen als ein vom Zufall abhängendes Ereignis aufgefaßt werden.

⁷⁵ Vgl. Eisenhart (1983, S. 534), Schneider (1988, S. 216).

einer diskreten Dreiecksverteilung. Hieran untersuchte er die Wahrscheinlichkeit von Abweichungen von Durchschnitten für eine symmetrische Verteilung um Null, die proportional zu 1, 2, 3, 4, 5, 6, 5, 4, 3, 2, 1 war. Zwei Jahre später konkretisierte er seine Forderungen bezüglich der Verteilung der Fehler, damit eine Wahl des Mittelwertes gerechtfertigt war. Zum einen sollten Abweichungen positiver und negativer Art sich ausgleichen, zum anderen sollten alle Fehler innerhalb vorgegebener Grenzen liegen, die von der Qualität des Instruments und der Geschicklichkeit des Beobachters abhingen.⁷⁶ Nun ging er von einer stetigen Dreiecksverteilung aus und leitete daraus die Verteilung des Mittelwertes dieser Fehlerverteilung ab. Unmittelbar darauf knüpften J.-L. Lagrange, J. H. Lambert, D. Bernoulli und P. S. Laplace an die Arbeiten Simpsons an und formulierten eine Reihe weiterer kontinuierlicher Fehlerverteilungen.⁷⁷ Anders als seine Vorgänger ging Laplace jedoch von einer *inversen* Wahrscheinlichkeitskonzeption aus und setzte sich damit explizit von seinen Vorgängern ab:

“The problem we are concerned with may be regarded from two different points of view, depending on whether we consider the observations before or after they are made. In the first case, the search for the mean we should take among the observations consists in determining *a priori* which function of the results of the observations is most advantageously taken for the mean result. In the second case, the search for this mean consists in determining a similar function *a posteriori*, that is, paying regard to the distances between the respective observations. We see easily that these two manners of regarding the problem must lead to different results but it is at the same time clear that the second is the only one that should be used. Nevertheless, it is from the first of these viewpoints that the question has, prior to now, been treated by Geometers. Thus however ingenious their researches have been, they can only be of very little use to observers.”⁷⁸

Wie sah nun die Wahrscheinlichkeits- und Inferenzkonzeption von Laplace aus? Zunächst schloß er den Zufall wie Bernoulli als etwas objektiv Gegebenes aus. “Zufall ist Ausdruck eines (vorläufigen) Informationsmangels über die wirkenden Ursachen, und Wahrscheinlichkeit ist demgemäß ein Maß für den subjektiven Informationsstand. Die Wahrscheinlichkeitsrechnung ist nach Laplace also gerade jene Disziplin, die überall dort, wo man vollständige kausale Erklärungen der Zusammenhänge noch nicht kennt, die beste Orientierungshilfe liefert.”⁷⁹ Laplace verstand die Begriffe Zufall und Wahrscheinlichkeit wie folgt:

“Wir betrachten eine Sache als Wirkung des Zufalls, wenn sie unseren Augen keinerlei Regelmäßigkeit oder Anzeichen einer Bestimmung bietet und wir außerdem die Ursachen, die sie hervorgebracht haben, nicht kennen. Der Zufall trägt also in sich keinerlei Wirklichkeitsanspruch; er ist nichts anderes als ein Begriff zur Beschreibung unserer Unwissenheit über die Art

⁷⁶ Stigler (1986, S. 95). Nach Stiglers Ansicht erfolgte dies wahrscheinlich durch einen kritischen Hinweis von T. Bayes.

⁷⁷ Zu Lagrange, Lambert und D. Bernoulli ausführlich Eisenhart (1983, S. 535-544).

⁷⁸ Pierre Simon Laplace, *Recherches sur le milieu qu'il faut choisir enter les résultats de plusieurs observations*, 1777, zitiert nach Stigler (1986, S. 119), Hervorhebungen im Original.

⁷⁹ Schneider (1988, S. 49f). Eine philosophische Fundierung erhielt eine solche Sichtweise etwa zur gleichen Zeit durch I. Kants *Kritik der reinen Vernunft*. Vgl. dazu Krüger (1987a, S. 61ff).

und Weise, in der sich die verschiedenen Teile einer Erscheinung miteinander und mit dem Rest des Kosmos miteinander verbinden. Der Begriff der Wahrscheinlichkeit bezieht sich auf diese Unwissenheit. [...] Die Wahrscheinlichkeit für die Existenz eines Ereignisses ist also nichts anderes als das Verhältnis der Anzahl der günstigen Fälle zu der aller möglichen Fälle, wenn wir außerdem keinen Grund sehen, weswegen einer dieser Fälle leichter eintreite als ein anderer. Sie kann folglich dargestellt werden durch einen Bruch, dessen Zähler die Anzahl der günstigen Fälle ist und dessen Nenner die aller möglichen Fälle.“⁸⁰

Die epistemologische Implikation dieser Wahrscheinlichkeitsdefinition war:

“If an event can be produced by a number n of different causes, then the probabilities of these causes given the event are to each other as the probabilities of the event given the causes, and the probability of the existence of each of these is equal to the probability of the event given that cause, divided by the sum of all the probabilities of the event given each of these causes.”⁸¹

In formaler Notation ist dies die berühmte Formel

$$(4) \quad P(A_i|E) = \frac{P(E|A_i)P(A_i)}{\sum_{j=1}^n P(E|A_j)P(A_j)}$$

wobei A_1, A_2, \dots, A_n die n Ursachen und E das Ereignis bezeichnen.⁸² Laplace ging zunächst davon aus, daß a priori jeder Ursache die gleiche Wahrscheinlichkeit $P(A_i) = 1/n$ zuzuordnen sei. Diese Vorstellung wurde später unter der Bezeichnung “Prinzip vom unzureichenden Grund” bekannt. Dann vereinfacht sich obige Formel zu

$$(5) \quad P(A_i|E) = \frac{P(E|A_i)}{\sum_{j=1}^n P(E|A_j)}$$

Nach Stigler handelt es sich bei der Annahme gleichwahrscheinlicher Ursachen nicht um eine metaphysische Überzeugung, sondern sie wurde von Laplace lediglich aus rechentechnischen Gründen gewählt. Da für Laplace A-priori-Wahrscheinlichkeiten Aussagen über unseren Wissensstand darstellten (s. o.), hielt er es für

⁸⁰ P. S. Laplace, *Recherches sur l'intégration des équations différentielles aux différences finies et sur leur usage dans la théorie des hasards*, in: *Mémoires de l'Académie Royale des sciences de Paris (Savants étrangers)* 7 (1773) 1776, zitiert nach der Übersetzung in Schneider (1988, S. 70).

⁸¹ S. P. Laplace, *Mémoire sur la probabilité des causes par les évènements*, in: *Mémoires de l'Académie royale des sciences présentés par divers savans* 6 (1774), S. 621-656, zitiert nach Stigler (1986, S. 102).

⁸² Vgl. ebda., S. 102. Verbal ausgedrückt: Die Wahrscheinlichkeit der Ursache A_i , wenn potentielle Wirkung E beobachtet wurde, ist gleich der Wahrscheinlichkeit der Wirkung E , wenn Ursache A_i vorliegt, mal der Wahrscheinlichkeit der Ursache A_i unabhängig von irgendeiner Beobachtung, dividiert durch die Summe der Wahrscheinlichkeiten aller Entstehungsprozesse von E aus sämtlichen möglichen Ursachen A_j . Von Randow (1992, S. 122f).

möglich, unterschiedliche Gewichte der einzelnen Ursachen durch eine feinere Aufteilung der "größeren" Ursachen wiederum in eine Gleichverteilung zu überführen.⁸³

Nachdem Laplace anhand der Fehlerrechnung zu seiner Konzeption der inversen Wahrscheinlichkeit gelangt war, konnte er dieses Prinzip allgemein anwenden. Laplaces Wahrscheinlichkeitskonzeption wurde im 19. Jahrhundert im Rahmen der Statistik allgemein akzeptiert.⁸⁴ Dennoch ist die Konzeption einer inversen Wahrscheinlichkeit nicht mit seinem Namen, sondern mit einer Arbeit von T. Bayes verbunden, und zwar mit einem posthum veröffentlichten Beitrag.⁸⁵ Über die Interpretation des Bayesschen Beitrages ist viel gestritten worden;⁸⁶ wir wollen uns daher auf eine kurze Zusammenfassung des Inhalts beschränken.⁸⁷ Bayes begann seine Abhandlung mit dem zu lösenden Problem:

"Gegeben die Anzahl Male, die ein unbekanntes Ereignis eingetreten und ausgeblieben ist. Gesucht die Chance, daß die Wahrscheinlichkeit seines Eintretens bei einem einzelnen Versuch irgendwo zwischen zwei angebbaren Graden der Wahrscheinlichkeit liegt."⁸⁸

Es folgen sieben Definitionen und sieben Lehrsätze, die die zeitgenössischen Ansichten zur Wahrscheinlichkeitstheorie zusammenfassen sollten. Von Interesse ist dabei seine Definition der Wahrscheinlichkeit:

"Die Wahrscheinlichkeit eines Ereignisses ist das Verhältnis zwischen dem Werte, welcher einer an das Eintreten des Ereignisses geknüpften Erwartung zu geben ist, und dem Werte des in diesem Falle erwarteten Gewinnes."⁸⁹

⁸³ Vgl. Stigler (1986, S. 103).

⁸⁴ Vgl. etwa Schneider (1987).

⁸⁵ Thomas Bayes [1702-1761] wurde als Sohn eines Geistlichen, einer der ersten sechs Nonkonformisten, die wieder öffentlich ordiniert wurden, geboren. T. Bayes war zunächst Assistent seines Vaters, später Geistlicher an zwei presbyterianischen Kirchen. 1742 wurde er aufgrund einer Arbeit zur Infinitesimalrechnung, wie bereits sein Vater, *Fellow* der *Royal Society*. Seine Schriften befaßten sich mit geistlichen wie auch naturwissenschaftlichen Problemen. Nach seinem Tod entdeckte der Pfarrer Richard Price, der Bayes' Nachlaß verwalten sollte, unter seinen nachgelassenen Schriften zwei Abhandlungen, die er zur Veröffentlichung empfahl. Eine davon behandelte asymptotische Reihen, die andere Probleme der Wahrscheinlichkeitsrechnung. Letztere wurde 1763 und 1764 in den Bänden 53 und 54 der *Philosophical Transactions* in einer von Price bearbeiteten Fassung unter dem Titel "An Essay towards solving a Problem in the Doctrine of Chances" veröffentlicht.

⁸⁶ Nach Stigler (1986, S. 127) unterlagen auch so bedeutende Wissenschaftler wie K. Pearson, H. Jeffreys, R. A. Fisher oder I. Hacking einer Fehlinterpretation des Bayesschen Ansatzes. Andere machen wiederum Laplace für Fehlinterpretationen verantwortlich und sehen Bayes' Darstellung als widerspruchsfrei an. Vgl. allgemein Stigler (1986, S. 98), Trader (1982), Menges (1967b) und Dale (1991, hier vor allem S. 1-50). Letztere Arbeit bietet eine enzyklopädische Darstellung der Behandlung inverser Wahrscheinlichkeit von T. Bayes bis K. Pearson.

⁸⁷ Wir verwenden die deutsche Übersetzung von Timerding, die in Schneider (1988, S. 108-115, 135-144) wiedergegeben ist.

⁸⁸ Thomas Bayes, *An essay towards solving a problem in the doctrine of chances*, zitiert nach Schneider (1988, S. 108).

⁸⁹ Ebda.

Die Wahrscheinlichkeit wird hier definiert als das Verhältnis der mathematischen Erwartung zum erwarteten Gewinn.⁹⁰ Weiter von Interesse ist der fünfte Lehrsatz, der die bedingte Wahrscheinlichkeit definiert:

“Wenn von zwei zusammenhängenden Ereignissen die Wahrscheinlichkeit für das zweite b/N und die Wahrscheinlichkeit für beide zusammen P/N ist, und man entdeckt, daß das zweite Ereignis wirklich eingetreten ist, so ist für die Richtigkeit der Vermutung, daß das erste Ereignis ebenfalls eingetreten ist, die Wahrscheinlichkeit P/b .”⁹¹

Mit diesen Bausteinen begann er die Lösung seines Problems. Bayes verwendete dazu eine geometrische Darstellungsweise.⁹² Er ging von einem (“Billard”-) Tisch (rechteckig, Seitenlänge a) aus, auf dem eine Kugel gerollt wird.⁹³ Der Punkt, an dem die Kugel zum Stillstand kommt, definiert einen Abstand x .⁹⁴

Nachdem die Position der ersten Kugel diese Flächen definiert hat, wird eine zweite Kugel n mal auf den Tisch gerollt. Wie groß ist nun die Wahrscheinlichkeit M , daß die Kugeln in der Region $E F D A$ liegen bleiben? Diese Wahrscheinlichkeit $p(M) = \theta$ ist keine feste Größe, sondern eine über dem Intervall $[0,1]$ gleichverteilte Zufallsvariable. Im 9. Lehrsatz gab Bayes an, wie die A-posteriori-Verteilung für θ unter der Bedingung des ersten Ereignisses bestimmt werden kann.⁹⁵

⁹⁰ Vgl. Menges (1967b, S. 488). Dies ist nach Menges *keine* subjektivistische Interpretation. Anders dagegen Stigler (1986, S. 124).

⁹¹ Thomas Bayes, An essay, zitiert nach Schneider (1988, S. 113). Diese Formulierung deutet unseres Erachtens auf eine subjektivistische Wahrscheinlichkeitsinterpretation hin.

⁹² Dadurch wirkt der Beweisgang sehr umständlich. Auch Stigler (1986, S. 123) betont, “Bayes’s essay [...] is extremely difficulty to read today – even when we know what to look for.” Die meisten Zeitgenossen von Bayes bedienten sich bereits der moderneren analytischen Darstellungsweise. Nach ebda., S. 124f, ermöglicht gerade diese Darstellungsweise einen Argumentationsgang, der mit einem Urnenmodell eines Bernoulli nicht möglich gewesen wäre.

⁹³ Vgl. zum folgenden auch Menges (1967b, S. 491f).

⁹⁴ Timerding (der deutsche Übersetzer) weist an dieser Stelle in einem Kommentar zu seiner Übersetzung zu Recht darauf hin, daß man fragen könne, warum überhaupt der Abstand x durch das Rollen einer Kugel realisiert und nicht einfach willkürlich bestimmt wird. Bayes war jedoch daran gelegen, die Wahrscheinlichkeit eines Ereignisses zu bestimmen, das wiederum durch ein anderes Ereignis bestimmt wurde. Vgl. dazu seine instruktiven Erläuterungen in Schneider (1988, S. 136, Anm. 109). Menges (1967b, S. 492): “Für Bayes symbolisiert das Rollen der ersten Kugel die ‘begleitenden Umstände’. Diese will er aus dem Ergebnis des Rollens der zweiten Kugel heraus erschließen.”

⁹⁵ Siehe Bayes in Schneider (1988, S. 141).

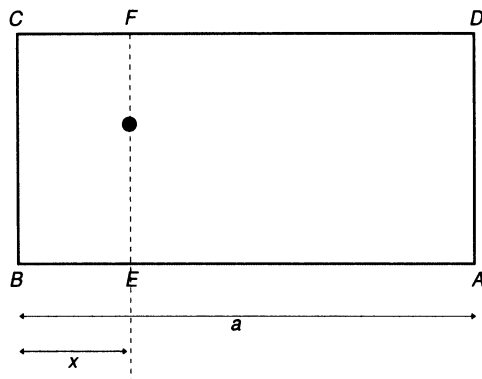


Abb.1: Bayes' "Billardtisch". Nach Menges (1967b, S. 491).

Der eigentlich umstrittene Punkt ist das nun folgende, dem 10. Lehrsatz vorausgehende "Scholium". Nach dem vorhergehenden Lehrsatz sei klar, daß man die Wahrscheinlichkeit des Ereignisses M , θ , aus der relativen Häufigkeit seines Eintretens und der Berechnung der entsprechenden Flächen bei mehreren Versuchen ermitteln könne. Entscheidend sei, daß man darüber hinausgehen könne.⁹⁶

"Daß dieselbe Regel auch geeignet ist, um in dem Falle irgendwelchen Ereignisses zu dienen, über dessen Wahrscheinlichkeit wir von vornherein durchaus nichts wissen, scheint aus der folgenden Überlegung hervorzugehen. Betreffs eines solchen Ereignisses habe ich keinen Grund zu vermuten, daß es bei einer bestimmten Anzahl von Versuchen eher eine gewisse Anzahl Male als eine andere eintreten würde. Denn ich kann in dieser Beziehung genau so überlegen, als ob die Wahrscheinlichkeit zuerst unbestimmt gewesen wäre und dann so bestimmt worden wäre, daß kein Anlaß zu der Annahme vorliegt, bei einer gewissen Anzahl von Versuchen werde das Ereignis irgend eine mögliche Anzahl eher als eine andere eintreten. Dies ist aber genau der Fall bei dem Ereignis M . [...] Im folgenden werde ich deswegen als zugegeben ansehen, daß die in betreff des Ereignisses M in Lehrsatz 9 gegebene Regel auch die Regel ist, die in Beziehung auf irgend ein Ereignis zu verwenden ist, von dessen Wahrscheinlichkeit vor irgendwelchen diesbezüglichen Versuchen oder Beobachtungen schlechterdings nichts bekannt ist. Und ein solches Ereignis werde ich unbekanntes Ereignis nennen."⁹⁷

Die späteren Interpreten des Scholiums haben darauf hingewiesen, daß hier zwar eine ähnliche Formulierung des "Prinzips vom unzureichenden Grund" gegeben wurde, wie wir sie schon bei Laplace fanden, allerdings die Annahme der Gleichverteilung in der bayesschen Konstruktion "physikalisch" gerechtfertigt war. Nicht sicher ist allerdings, ob Bayes die Verwendung einer Gleichverteilung wie Laplace auch dann vorsah, wenn sich dies aus dem Problem nicht zwingend ergab.⁹⁸

⁹⁶ Wir wollen diesen Abschnitt wörtlich zitieren, der nach Menges (1967b, S. 492) "schon schwer genug zu lesen ist".

⁹⁷ Thomay Bayes, An essay, zitiert nach Schneider (1988, S. 142f). Hervorhebung von uns.

⁹⁸ Menges (1967b, S. 493): "Die Art und Weise wie Bayes sich ausdrückt, läßt vermuten, daß er eine Neigung dazu verspürte, aber mit Sicherheit läßt es sich nicht behaupten."

Die eigentliche Verbreitung der bayesschen Konzeption erfolgte durch Laplace.¹⁰⁰ Der von Laplace geschaffene Wahrscheinlichkeitsapparat, der für die Binomialverteilung voll und hinsichtlich der Fehlertheorie ansatzweise entwickelt war, konnte nun mit der durch Legendre 1805 vorerst abgeschlossenen Entwicklung der Kleinstquadratmethode verbunden werden. Bis zu dieser Zeit beruhte die Überzeugung, daß dieses Verfahren die beste Lösung für ein Meßproblem darstellte, mehr oder weniger auf der Intuition; eine Quantifizierung der jeweiligen Unsicherheit, eine Antwort auf die Frage "How good is best"¹⁰¹ war jedoch nicht möglich. Die Synthese von Wahrscheinlichkeitstheorie und Kleinstquadratverfahren wurde in den nächsten Jahren vor allem von Laplace und Carl Friedrich Gauss [1777-1855] vollzogen. Bahnbrechende Wirkung sollte dabei eine Arbeit von Gauss aus dem Jahr 1809 haben, in der er erstmals die Fehlerrechnung wahrscheinlichkeitstheoretisch einbettete. Gauss widmete sich dem Problem der "Kombination von Beobachtungen" in einer Arbeit über Planetenbewegungen.¹⁰² Im Unterschied zu Mayer, Boscovich, Laplace oder Legendre konstruierte er jedoch von vornherein einen wahrscheinlichkeitstheoretischen Rahmen. Legendre hatte bereits aus intuitiven Überlegungen heraus eine Methode der kleinsten Quadrate entworfen (s. o.); Gauss war jedoch daran gelegen, diese Wahl als die aus wahrscheinlichkeitstheoretischen Gesichtspunkten beste zu begründen. Das Problem, das sich stellte, war wiederum die Ermittlung der Koeffizienten $p, q, r, s \dots$ eines Gleichungssystems

$$(6) \quad \begin{array}{l} y_1 = p \cdot x_{1,1} + q \cdot x_{2,1} + r \cdot x_{3,1} + s \cdot x_{4,1} + \cdots \\ y_2 = p \cdot x_{1,2} + q \cdot x_{2,2} + r \cdot x_{3,2} + s \cdot x_{4,2} + \cdots \\ \vdots = \vdots \quad \vdots \end{array}$$

falls die Zahl der Gleichungen die Zahl der Unbekannten überstieg.¹⁰³ Die Koeffizienten könnten durch Beobachtungen z_1, z_2, z_3 ermittelt werden. Bei der Lösung dieses Problems folgte Gauss Laplace und verwendete hierzu dessen Form des "Bayes-" Theorems.¹⁰⁴ Er ging davon aus, daß allen Werten für die Unbekannten *a priori* die gleiche Wahrscheinlichkeit zukam. Die gesuchten Koeffizienten waren

⁹⁹ Auf rechentechnische Einzelheiten wollen wir hier verzichten.

¹⁰⁰ Stigler (1986, S. 100): "In fact, Bayes's failure to publish his own paper may well have been due more to his own view of the work as analytically incomplete and of little or no practical use than to any philosophical misgivings. Laplace was able to overcome that difficulty." Dale (1991, S. 99) gelangt zu der Überzeugung: "Perhaps one should consider Laplace, rather than Bayes, as the father (albeit unwittingly) of inverse probability. Auch Porter (1986, S. 93) vertritt diese Ansicht.

¹⁰¹ Stigler (1986, S. 140).

¹⁰² Carl Friedrich Gauss, *Theoria motus corporum coelestium in sectionibus conicis Solem ambientium*, Hamburg 1809. Die entscheidenden Abschnitte finden sich in deutscher Übersetzung in Schneider (1988, S. 237-243).

¹⁰³ Wir verwenden der Einfachheit halber eine modernere Notation. Die y -Werte sind keine Beobachtungen, sondern die exakten Werte.

¹⁰⁴ Vgl. Stigler (1986, S. 14).

die Werte, denen *nach* der Beobachtung die größte *A-posteriori*-Wahrscheinlichkeit zukam. Dies waren die Werte, die das Produkt der Fehlerverteilungen maximierten und die durch partielle Ableitungen bestimmt werden konnten.¹⁰⁵ Die Größen der möglichen Fehler $\Delta_1 = y_1 - z_1$, $\Delta_2 = y_2 - z_2$, $\Delta_3 = y_3 - z_3$, ... waren durch Wahrscheinlichkeitsverteilungen $\varphi(\Delta_1)$, $\varphi(\Delta_2)$, $\varphi(\Delta_3)$, ... bestimmt, deren Form sich aus Plausibilitätsbetrachtungen ergab. Im Gegensatz zu Laplace wurde die Argumentationskette aber bei Gauss auf den Kopf gestellt. Man könne lediglich allgemeine Aussagen über die Fehlerverteilung treffen:¹⁰⁶

“Wir nehmen zuerst an, es sei bei allen Beobachtungen die Sachlage richtig gewesen, dass kein Grund vorhanden ist, die eine für weniger genau als die andere zu erachten, oder dass man gleich grosse Fehler bei den einzelnen für gleich wahrscheinlich halten muss.¹⁰⁷ Die Wahrscheinlichkeit, welche irgend einem Fehler Δ beizulegen ist, wird daher durch eine Funktion von Δ ausgedrückt, welche wir mit $\varphi(\Delta)$ bezeichnen wollen. Wenn man nun auch diese Funktion nicht genau angeben kann, so kann man doch wenigstens versichern, dass ihr Werth ein Maximum für $\Delta = 0$ werden müsse, dass er im allgemeinen für gleiche und entgegengesetzte Werte von Δ der gleiche sei, und endlich, dass er verschwinde, wenn man für Δ den grössten Fehler oder noch einen grösseren Werth annimmt. Eigentlich ist deshalb $\varphi(\Delta)$ zu der Gattung der unstetigen Funktionen zu rechnen,¹⁰⁸ und wenn wir uns erlauben, zum praktischen Gebrauch an ihrer Stelle eine analytische Funktion einzuführen, so muss diese so beschaffen sein, daß sie von $\Delta = 0$ ab nach beiden Seiten gleichsam asymptotisch gegen 0 convergirt, so dass sie ausserhalb der betreffenden Grenze als wirklich verschwindend angesehen werden kann.”¹⁰⁹

Die Annahme einer Normalverteilung ergab sich also keineswegs zwingend, etwa als Grenzverteilung von Summen von Elementarfehlern, sondern aus rein pragmatischen Gründen, weil sie mathematisch leichter handhabbar war.¹¹⁰ Gauss sah sich sogar gezwungen, die Tatsache, daß der wahrscheinlichste (Maximum-Likelihood) Wert bei mehreren Beobachtungen unter gleichen Bedingungen das arithmetische Mittel ist, als “Axiom” einzuführen:

“Da diese [die Funktion der Fehlerverteilung, T. R.] aber a priori nicht definiert werden kann, so wollen wir die Sache von einer anderen Sache angreifen, und nachforschen, auf welcher stillschweigend gleichsam als Grundlage angenommenen Funktion ein landläufiges Princip eigentlich beruht, dessen Vortrefflichkeit allgemein anerkannt ist. Wie ein Axiom pflegt man nämlich die Hypothese zu behandeln, wenn irgend eine Grösse durch mehrere unmittelbare, unter gleichen Umständen und mit gleicher Sorgfalt angestellte Beobachtungen bestimmt worden ist, dass alsdann das arithmetische Mittel zwischen allen beobachteten Werthen, wenn

¹⁰⁵ Dies ist das mittlerweile klassische Maximum-Likelihood-Argument.

¹⁰⁶ Stigler (1986, S. 141): “Instead of imposing further conditions directly, he assumed the conclusion!”.

¹⁰⁷ Dadurch reduzieren sich die n Fehlerverteilungen zu einer einzigen.

¹⁰⁸ Gemeint ist hier z. B. entweder eine Dreiecksverteilung oder (für nicht völlig verschwindende Fehler) eine Doppel exponentialverteilung.

¹⁰⁹ Carl Friedrich Gauss, *Theoria motus*, zitiert nach Schneider (1988, S. 237f).

¹¹⁰ Dennoch wird sie im folgenden zu einer universalen Gesetzmäßigkeit hochstilisiert, so daß um die Jahrhundertwende der französische Physiker G. Lippman zu H. Poincaré bemerkte: “All the world believes it firmly, because the mathematicians imagine that it is a fact of observation, and the observers that it is a theorem of mathematics.” Zitiert nach Eisenhart (1983, S. 531).

auch nicht mit absoluter Strenge, so doch wenigstens sehr nahe den wahrscheinlichsten Werth gebe, so dass es immer das sicherste ist, an diesem festzuhalten.“¹¹¹

Unter dieser Voraussetzung, konnte er nachweisen, werde das Produkt der Fehlerverteilungen nur maximiert, wenn diese Fehlerverteilungen die Form der Normalverteilung

$$(7) \quad \varphi(\Delta) = \frac{h}{\sqrt{\pi}} e^{-h^2 \Delta^2}$$

aufweisen, wobei h als Maß für die Genauigkeit der Beobachtungen angesehen werden könne.¹¹² Eine Folgerung daraus sei, daß das Produkt der Fehlerverteilungen nur dann maximal werde, wenn die Summe der quadrierten Abweichungen ein Minimum ist:

*“Das wahrscheinlichste Werthsystem der Unbekannten p, q, r, s etc. wird daher dasjenige sein, bei welchem die Quadrate der Differenzen zwischen den beobachteten und berechneten Werthen [...] die kleinste Summe ergeben, wenn nur bei allen Beobachtungen der gleiche Grad von Genauigkeit vorausgesetzt werden darf.”*¹¹³

Wie Stigler zu Recht betont, ist die hier dargelegte Argumentationskette rückblickend betrachtet fehlschlüssig und zirkulär. Auch Gauss sah diese Schwäche, als er sich über ein Jahrzehnt später nochmals dieser Problematik widmete. Entscheidend war jedoch, daß unmittelbar nach der Publikation der Abhandlung Laplace darauf aufmerksam wurde und ihm plötzlich die Verbindung zwischen dem zentralen Grenzwertsatz und der Koeffizientenbestimmung klar wurde.¹¹⁴ Laplace argumentierte: Wenn die Fehler in den Gleichungssystemen aggregierte Größen darstellten, dann seien sie aufgrund des Zentralen Grenzwertsatzes normal- oder “Gauss-”verteilt.¹¹⁵ Wenn diese Form von vornherein gegeben war, dann ergab sich die Optimalität (im Sinne der Likelihoodfunktion, nach damaliger Interpretation unter Annahme einer A-priori-Gleichverteilung das Maximum der A-posteriori-Verteilung) des Kleinstquadratverfahrens. Damit waren auch die Bedenken von Gauss, daß es sich bei der von ihm vorgeschlagenen Funktion nicht um die exakte Fehlerverteilung handelte, haltlos. Diese Erkenntnis war aus konzeptueller Sicht von enormer Bedeutung: “The remarkable circumstance that the curve that led to the simplest analysis also had such an attractive rationale was conceptually liberating.”¹¹⁶ Zwei über hundert Jahre parallel verlaufende Entwicklungen hatten hier eine Synthese gefunden.

¹¹¹ Carl Friedrich Gauss, *Theoria motus*, zitiert nach Schneider (1988, S. 241).

¹¹² Gauss betonte, daß diese Funktion keine exakte Fehlerverteilung darstellen könne, “da die möglichen Fehler immer in bestimmten Grenzen eingeschlossen sind”. Dieser Mangel sei jedoch für alle praktischen Zwecke ohne Bedeutung.

¹¹³ Carl Friedrich Gauss, *Theoria motus*, zitiert nach Schneider (1988, S. 243). Hervorhebung im Original. Die Entdeckung des Kleinstquadratverfahrens nahm Gauss zwar auch für sich in Anspruch; mittlerweile gilt jedoch als gesichert, daß Legendre die Urheberschaft gebührt. Gauss hat im folgenden einige Rechenvereinfachungen entwickelt.

¹¹⁴ Vgl. Stigler (1986, S. 143).

¹¹⁵ Dies wird allgemein als Elementarfehlerhypothese bezeichnet.

Nach Stigler (1986, S. 157) markiert der Tod Laplaces 1827 das Ende einer Ära. Zu diesem Zeitpunkt hatte Gauss eine Erweiterung der Laplaceschen Theorie veröffentlicht, die das zirkuläre Argument von 1809 überwunden hatte,¹¹⁷ und auch andere Mathematiker, Astronomen und Geophysiker bauten dieses Gebiet immer weiter aus.¹¹⁸ Bis zur Mitte des 19. Jahrhunderts entwickelte sich die Gauss-Laplace-Synthese zur Standardsicht in bezug auf die Themen "Wahrscheinlichkeit" und "Kleinstquadrate".¹¹⁹ Nahezu alle Beiträge zu diesem Thema (Monographien, Lexikonartikel) bauten auf ihrer Synthese auf oder erweiterten sie. Dennoch war diese Synthese lange Zeit auf einen engen Anwendungsbereich begrenzt:

"In many respects it was one of the major success stories in the history of science. Yet it also poses a puzzle, for the applications of this marvelous statistical technology were widespread only geographically; to an amazing degree they remained confined to the narrow set of disciplines that spawned them. They became commonplace in astronomical and geodetic work while remaining almost unknown in the social sciences, even though their potential usefulness in the social sciences was known from the time of Bernoulli and even though some early social scientists (for example Adolphe Quetelet) were also astronomers."¹²⁰

Vor allem konzeptionelle Barrieren verhinderten, daß die Techniken in die Sozialwissenschaften übernommen wurden. Das Hauptproblem der Übertragung statistischer Methodologien aus dem Anwendungsbereich der Astronomie und Geodäsie in die Sozialwissenschaften wurde die Bildung von Gruppen, innerhalb derer Variationen als zufällig angenommen werden konnten und große bedeutende Einflußfaktoren berücksichtigt wurden.¹²¹

Die Adaption dieser Konzepte in den Sozialwissenschaften ist untrennbar mit dem Namen Adolphe Quetelets verknüpft, und zwar zunächst mit einem Mißerfolg.¹²²

¹¹⁶ Stigler (1986, S. 145).

¹¹⁷ Carl Friedrich Gauss, *Theoria combinationis observationum erroribus minimis obnoxiae (pars prior)*, in: *Commentationes societatis regiae scientiarum Göttingensis recentioris* 5 (1823). Die entscheidenden Abschnitte finden sich in deutscher Übersetzung in Schneider (1988, S. 244-257).

¹¹⁸ Auf diese Weiterentwicklungen wollen wir hier nicht weiter eingehen. Einige Texte finden sich in ebda., Kap. 6. Vgl. auch Harter (1983) und Eisenhart (1983).

¹¹⁹ Stigler (1986, S. 158): "The Gauss-Laplace synthesis brought together two well-developed lines – one of the combination of observations through the aggregation of linearized equations of condition, the other the use of mathematical probability to assess uncertainty and make inferences – into a coherent whole."

¹²⁰ Stigler (1986, S. 158).

¹²¹ In der Astronomie und Geodäsie existierten eindeutige Theorien in Form von funktionalen Zusammenhängen. Abweichungen konnten zumeist ohne konzeptionelle Schwierigkeiten als Meßfehler angesehen werden. In bezug auf soziale Phänomene ist jedoch die Messung selbst Gegenstand des Interesses, und die Stärke der Variation hängt maßgeblich von vorher durchgeführten Klassifikationen und Einteilungen ab.

¹²² A. Quetelet [1796-1874] wurde in Belgien geboren, promovierte 1819 an der Universität Gent mit einer Arbeit über konische Schnitte und lehrte seit diesem Jahr Mathematik am Athenäum in Brüssel. 1823/24 ging er nach Paris, wo er bei J. Fourier und möglicherweise auch bei Laplace die Wahrscheinlichkeitstheorie erlernte. 1833 baute er im Auftrag der belgischen Regierung ein Observatorium. Quetelet schrieb etwa ein Dutzend Bücher und zahlreiche Aufsätze zu den verschiedensten Themen und war auch einer der einflußreichsten Wissenschaftspolitiker seiner Zeit.

Ein erster Ansatz, wahrscheinlichkeitstheoretische Erwägungen auf gesellschaftliche Daten anzuwenden, war sein Vorschlag, einen Zensus nach einer Methode durchzuführen, die Laplace für Frankreich vorgeschlagen hatte. Diese Methode bestand darin, anhand der Anzahl der Geburten der Geburtsregister oder der Todesfälle der Sterberegister, multipliziert mit dem *geschätzten* Verhältnis von Gesamtbevölkerung und Geburten/Todesfälle, die Anzahl der Gesamtbevölkerung zu ermitteln. Das Verhältnis sollte dabei aufgrund einiger sorgfältig ausgewählter Gemeinden bestimmt werden.¹²³ Nach ersten Berechnungen, die Quetelet auf diese Weise 1824 für Belgien anstellte, entschied er sich jedoch überraschenderweise für die Empfehlung einer Totalerhebung.¹²⁴ Diese Kehrtwendung erfolgte nach Ansicht Stiglers durch das sog. Keverberg-Argument.¹²⁵ Keverberg sah es als unmöglich an, durch das von Laplace entwickelte und von Quetelet vorgeschlagene Vorgehen zu präzisen Resultaten zu gelangen. Sein Hauptargument war die überwältigende Anzahl möglicher Umstände, die eine Auswirkung auf die Mortalitäts- und Geburtsraten haben könnten. Diese unterschieden sich in Städten oder auf dem flachen Land, für große reiche Städte und arme kleine Dörfer, bei unterschiedlicher Bevölkerungsdichte, variierender natürlicher Umgebung, unterschiedlicher Entfernung zum Meer, unterschiedlichen persönlichen Lebensumständen (Ernährung, Lebenswandel etc.) sowie einer Reihe weiterer lokaler Umstände, kurzum:

“It must therefore be extremely difficult, not to say impossible, to determine in advance with any precision, based on incomplete and speculative knowledge, the combination of all of these elements that in fact exist. There would seem to be infinite variety in the nature, the number, the degree of intensity, and the relative proportion of these elements.”¹²⁶

Um all diese Faktoren angemessen zu berücksichtigen, müsse eine Teilerhebung aus so vielen Elementen bestehen, daß gegenüber einer Totalerhebung nichts gewonnen wäre. Quetelet scheint sich spätestens dann diesem Argument gebeugt zu haben, als seine durchgeführten Berechnungen nach der Laplaceschen Methode enttäuschende Resultate erbrachten. Legte man die für Frankreich anhand der dreißig Departements ermittelten Geburtsraten zugrunde, errechnete sich für die Niederlande und Luxemburg eine Einwohnerzahl von 6.9242.424, anhand der Sterberaten dagegen von 5.351.628. Auch einige für belgische Regionen ermittelte Raten ergaben derart starke Unterschiede, daß sie das Argument Keverbergs zu bestätigen schienen. Quetelet war nun überzeugt, daß eine unbegrenzte Anzahl von Faktoren die Größe, die er messen wollte, beeinflusste und er nicht in der Lage war, diejenigen, die tatsächlich von Bedeutung waren, herauszufinden.¹²⁷

¹²³ Für Frankreich etwa dreißig, die über das Land gleich verteilt sein sollten, damit das Resultat unabhängig von lokalen Einflüssen sei.

¹²⁴ Vgl. Stigler (1986, S. 163ff). Im Gegensatz zu Lazarsfeld (1961), der wiederum auf Stephan (1948) verweist, sieht Stigler hierin nicht nur eine Diskontinuität in der methodischen Entwicklung durch die Sozialwissenschaften, sondern ist der Ansicht, daß dies eine “deeper conceptual explanation” erfordere als Ignoranz durch spätere Wissenschaftler.

¹²⁵ Baron de Keverberg war einer der für den Zensus zuständigen offiziellen Berater.

¹²⁶ Zitiert nach Stigler (1986, S. 165).

¹²⁷ Eine solche praktische Erfahrung ist Laplace erspart geblieben. Diese Erfahrung ist anfangs

Den Hauptbeitrag für die statistische Analyse sozialer Daten durch Quetelet sieht Stigler in zwei Ansätzen: zum einen durch sein Konzept des *l'homme moyen*, zum anderen durch seine Anpassung von Verteilungen. Zwischen 1827 und 1835 untersuchte Quetelet eine Vielzahl von (bivariaten) Beziehungen in Form von Tabellen oder Graphiken: Geburts- und Todesraten in Abhängigkeit von den Monaten und der Temperatur, den Zusammenhang von Mortalität und Berufen und Orten, in Gefängnissen und Krankenhäusern usw. Seine Erkenntnisse resultierten schließlich 1835 in einer ersten Buchausgabe seiner *Physique sociale*, die ihm internationale Beachtung als Sozialwissenschaftler einbrachte. Aufbauend auf Daten einer großen Anzahl französischer Soldaten ermittelte er z. B. deren durchschnittliche Körpergröße und ihr durchschnittliches Gewicht, Werte, die später einmal mit den entsprechenden Werten von Soldaten anderer Nationen verglichen werden sollten. Aber auch Kennzahlen aus dem Gebiet der "Moralstatistik" wurden ermittelt. So setzte er die für eine Gruppe errechnete Arrest- oder Trunkenheitsrate mit der Neigung ("propensity") einer Person dieser Gruppe, kriminell oder betrunken zu werden, gleich.¹²⁸ Diese Zahlen sollten einen Vergleichsmaßstab für verschiedene Gruppen oder z. B. die verschiedenen Geschlechter bieten. Quetelet behauptete nicht, einen universalen *homme moyen* ermittelt zu haben, sondern mehrere, deren konkrete Ausgestaltungen von bestimmten (zu benennenden) Faktoren abhingen: so gäbe es einen *homme moyen* für jedes Alter, für jede Rasse, jedes Land usw. In diesen durchschnittlichen Menschen seien alle zufälligen Variationen einer Gesellschaft beseitigt.¹²⁹ Stigler bemerkt hierzu:

"He was, like the astronomers before Laplace, unable to take the conceptual step of combining measurements taken under a variety of conditions into one analysis. He could write of marginal relationships, and within categories he could seek relations across other variables (such as male height across age); but he could not go further. His elimination of individual variation was at bottom reduced to an assumption of *ceteris paribus* – that all other things being equal."¹³⁰

Quetelets Ansatz hatte Auswirkungen auf verschiedene Richtungen,¹³¹ doch war für die Entwicklung statistischer Konzepte vor allem die Aufnahme seiner Ansätze durch F. Galton von Bedeutung.

auch in der Astronomie gemacht worden, bevor erfolgreiche Prognosen Vertrauen in die "Kombination von Beobachtungen" schufen.

¹²⁸ Vermutlich war auch ihm der Unterschied zwischen der erfaßten Arrestrate und dem Anteil krimineller Personen sowie der wegen Trunkenheit registrierten und tatsächlich "betrunkenen" Personen klar.

¹²⁹ Seine Sicht hatte sich zwischen 1835 und 1848 gewandelt. Während er anfangs noch an der gesamten Verteilung um den mittleren Wert (den er in Analogie zur Physik als Gravitationszentrum auffaßte) interessiert war, richtete sich später seine ganze Aufmerksamkeit auf die Mitte, um "Typen" zu kennzeichnen. Vgl. Stigler (1986, S. 172).

¹³⁰ Ebda., S. 173.

¹³¹ Etwa im Zusammenhang materialistischer geschichtsteologischer Deutungen von Karl Marx oder Henry Thomas Buckle. Buckles 1857 und 1861 erschienene zweibändige *History of Civilisation in England* fand sowohl in wissenschaftlichen als auch in populären Kreisen enorme Beachtung. Die beiden Bände erlebten zahlreiche Neuauflagen und Übersetzungen unter anderem in die französische, russische und deutsche Sprache. Vgl. dazu Porter (1986, S. 60ff).

Galton¹³² knüpfte unmittelbar an Quetelet an. In seiner ersten größeren Arbeit sah er in Anlehnung an Quetelet in menschlichen Körpergrößen ein Normalverteilungsgesetz verwirklicht. Dies könne sogar als Überprüfung dafür verwendet werden, ob die Beobachtungen, die in einer solchen Kurve zusammengefaßt wurden, einer homogenen Gruppe zugehören, da Personen, die unterschiedlichen Gruppen entstammten (z. B. unterschiedlichen Rassen), getrennt behandelt zwar zu Normalverteilungen führen würden, ihre gemeinsame Behandlung aber eine gemischte Verteilung produziere: "The law may, therefore, be used as a most trustworthy criterion, whether or not the events of which an average has been taken, are due to the same or to dissimilar classes of conditions."¹³³ Galton ging aber einen entscheidenden Schritt über Quetelet hinaus. Für ihn war das Bilden solcher homogenen Gruppen nicht das Ende, sondern der Anfang der Analyse. Er ging der folgenden Frage nach: Wenn sich aufgrund einer Vielzahl voneinander unabhängiger Faktoren, von denen keiner einen dominanten Einfluß ausübt, eine Normalverteilung z. B. für die Körpergröße herausstellt, wie ist dann der Einfluß eines einzelnen Faktors – z. B. der Elternschaft – zu beurteilen? "It is a tribute to Galton's own genius and to his persistent work over twenty years that he was able to discover a way around this impasse, with his formulation of regression and its link to the bivariate normal distribution."¹³⁴ Diese Synthese, die er 1889 in seinem berühmten Buch *Natural Inheritance* darlegte, hatte er seit 1874 über die folgenden Zwischenschritte entwickelt: 1. durch eine nähere Untersuchung von Fehlerursachen, 2. anhand von Experimenten mit Erbsen, die 1877 eine erste Fassung seines Regressionsansatzes brachten und 3. durch die Entwicklung eines mathematischen Rahmens für diese Regressionsrechnung anhand von Populationsdaten im Jahre 1885.

1. Zunächst suchte Galton eine Erklärung für folgende Situation: Die Laplacesche Fassung des Zentralen Grenzwertsatzes besagte, daß sich Größen einer Normalverteilung annähern, wenn eine Reihe von Bedingungen, wie Unabhängigkeit, gegenseitige Aufhebung und große Anzahl der Fehler, erfüllt waren. Nun waren aber in den von ihm betrachteten Daten ("vital and social phenomena") diese Annahmen keineswegs berechtigt, und dennoch bildeten sie eine Normalverteilungsform. Vor allem die Tatsache, daß dies trotz der Einwirkung einzelner größerer Faktoren geschah, weckte sein Interesse. Seine experimentellen Reihen mit Erbsen, die mit dem "aspect" einem Haupteinflußfaktor für ihre Größe ausgesetzt und dennoch normalverteilt waren, führten ihn zu einer Erkenntnis, die Stigler (1986, S. 281) als den "most important step in perhaps the single major breakthrough in statistics in the last half of the nineteenth century" bezeichnet. Diese Erkenntnis erzielte er nicht aufgrund einer mathematischen Überlegung, sondern auf einem

¹³² Francis Galton [1822-1911], ein Vetter Charles Darwins, wird von Stigler (1986, S. 266) als "perhaps the last of the gentlemen scientists" bezeichnet. Er studierte in Cambridge Medizin und war Zeitgenosse zunächst durch Berichte seiner Südwestafrika-Expeditionen bekannt. Er hatte nie eine akademische Position inne. Seit 1860 war er *Fellow* der *Royal Society*.

¹³³ Francis Galton, *Hereditary genius: An inquiry into its laws and consequences*, London 1869 [2. A. 1892], zitiert nach Stigler (1986, S. 268).

¹³⁴ Ebda., S. 272.

anschaulichen Wege. Analog zu seiner berühmten Konstruktion, die die Binomialverteilung illustrierte, dem "Galtonschen Brett", entwickelte er eine weitere Konstruktion, allerdings nur auf dem Papier. In dieser war der Schacht auf mittlerer Höhe durch eine Linie AB geteilt, auf der sich analog zur ersten Fassung verschiedene Abteilungen befanden. Falls die Anzahl der Nagelreihen oberhalb dieser Reihe genügend groß war, ergab sich hier eine approximativ normalverteilte Binomialverteilung. Nun wird eine dieser Abteilungen geöffnet, wodurch die Kugeln über eine analoge Nagelreihe wiederum in eine bestimmte Anzahl von Abteilungen fallen, abermals in Form einer Normalverteilung, die in Abhängigkeit von der Zahl der Kugeln, die sich in der Abteilung befand, unterschiedlich groß, aber von derselben Streuung sein wird. Wenn nun alle Abteilungen der Ebene AB geöffnet werden, entsteht eine Vermischung dieser Kurven verschiedener Größe. Da aber diese Situation gleichbedeutend ist mit dem ersten Galtonschen Brett (ohne Zwischenebene), muß diese gemischte Verteilung wieder eine Normalverteilung sein. Die Abteilungen der Ebene AB können als Ausprägungen eines "bedeuten-den" Faktors (der selbst wiederum das Ergebnis vieler kleiner ist / sein kann) angesehen werden, in Galtons Erbsenexperimenten z. B. die unterschiedlichen Ausprägungen des Faktors "aspect", die untere Ebene als Größe der Erbsen. Diese Anordnung brachte den anschaulichen Nachweis, daß Summen von Normalverteilungen wiederum normalverteilt sind. Zwar war diese Erkenntnis prinzipiell schon seit De Moivre und Laplace bekannt, aber nicht deren konzeptionelle Verwendung: "This concept freed Galton of the restrictions of naive error theory without requiring that he go beyond that theory. He could conceive of his data as a mixture of very different populations, notwithstanding the unity apparent in its normal outline."¹³⁵

2. Der zweite Schritt bestand in einem Erbsenexperiment, das Galton präsentierte. Er hatte sieben Freunden jeweils 70 Samenkörner gegeben, die nach der Größe geordnet jeweils in sieben Kategorien eingeteilt waren. Er fand heraus, daß die Nachkommen jeder Gruppe wiederum normalverteilt waren, wobei sie zwar unterschiedliche Mittelwerte, aber die gleiche Streuung aufwiesen.¹³⁶ Wie konnte es aber sein, daß die Streuung der Nachkommen nicht zu einer größeren Streuung insgesamt führte? Galtons Antwort – "his second major finding" – lautete *reversion*.¹³⁷ Zwar waren die Abkömmlinge der jeweiligen Gruppen normalverteilt mit gleicher Streuung, doch war ihr Mittelwert nicht mit dem Mittelwert ihrer Abstammungsgruppe identisch, sondern tendierte zum allgemeinen Mittelwert der elterlichen Verteilung. Die Abweichungen der Mittelwerte der Abkömmlinge vom Gesamtmittelwert gingen in dieselbe Richtung wie die Abweichung des Mittelwertes ihrer Elterngruppe, jedoch betrug die Größe dieser Abweichung lediglich ein Drittel von der Abweichung der elterlichen Gruppe. Dadurch wurde der wachsende

¹³⁵ Ebda., S. 281.

¹³⁶ Strenggenommen hätte er nicht von einer gleichen Anzahl zu Beginn ausgehen müssen, sondern von einer gleichen Anzahl, die sich schließlich auf der Ebene AB in seinem Beispiel befindet. Vgl. Stigler (1986, S. 282).

¹³⁷ Ebda., S. 282.

Streuungseffekt wieder ausgeglichen, und die Verteilung der Erbsen blieb stabil. Nun fehlte lediglich der dritte Schritt zu einem allgemeinen "Regressions"ansatz: eine Begründung für dieses Phänomen.

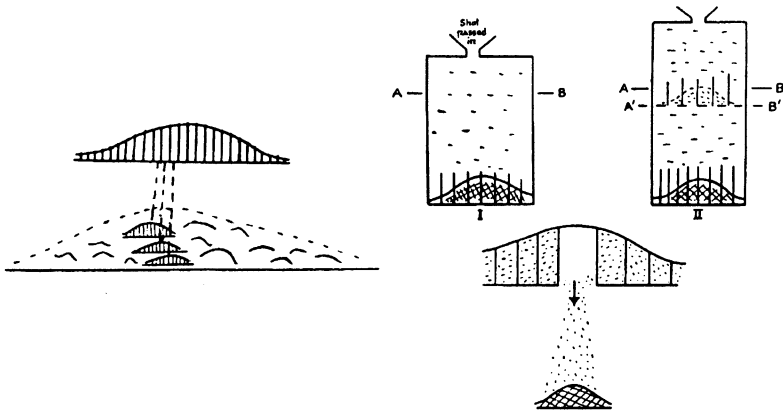


Abb. 2: Funktionsweise des zweistufigen Galtonschen Brettes. Zeichnung von Karl Pearson, angefertigt nach Vorlagen von Galton. Quelle: Stigler (1986, S. 279).

3. In der Folgezeit erweiterte Galton seine Datenbasis um anthropometrisches Material. Folgende Schwierigkeiten ergaben sich zunächst: Jede Person hatte zwei Elternteile, Frauen und Männer hatten jedoch eine unterschiedliche mittlere Größe. Dieses Problem umging er dadurch, daß er die Größe der Frauen mit 1.08 multiplizierte, eine Konstante, die sich aus seinem umfangreichen Datenmaterial ergab.¹³⁸ Doch welche Größe war nun entscheidend: die des Vaters oder die (mit dem Faktor 1.08 multiplizierte) der Mutter? Konnte man den Durchschnitt beider Größen verwenden? Waren die Größen der Kinder nur abhängig von der durchschnittlichen Größe der Eltern, oder spielte auch die Differenz ihrer Größen eine Rolle? Eine Überprüfung an 525 Kindern ergab, daß letzteres nicht der Fall war.

Die vorliegende Situation unterschied sich von dem vorhergehenden Erbsenexperiment in einem wichtigen Punkt: Dort untersuchte er Verteilungen in Abhängigkeit von sieben verschiedenen Größen der vererbenden Erbsen. Hier gab es keine solchen natürlichen Einteilungen; diese mußten von ihm selbst vorgenommen werden. Galton teilte die korrigierte Durchschnittsgröße der Eltern in 11, die Größe der Kinder in 14 Gruppen ein und bildete anhand der von ihm gemessenen 928 Fälle eine Kreuztabelle.

¹³⁸ Die Tatsache, daß die Größe des Partners ein Kriterium bei der Partnerwahl sein konnte, schloß er aus.

Tab. 1: Durchschnittsgröße von Eltern und Kindern nach Galton. Quelle: Stigler (1986, S. 286).

Height of the mid-parent in inches	Height of the adult child														Total no. of adult children	Total no. of mid-parents	Medians
	<61.7	62.2	63.2	64.2	65.2	66.2	67.2	68.2	69.2	70.2	71.2	72.2	73.2	>73.7			
>75.0	—	—	—	—	—	—	—	—	—	—	—	1	3	—	4	5	—
72.5	—	—	—	—	—	—	—	1	2	1	2	7	2	4	19	6	72.2
71.5	—	—	—	—	1	3	4	3	5	10	4	9	2	2	43	11	69.9
70.5	1	—	1	—	1	1	3	12	18	14	7	4	3	3	68	22	69.5
69.5	—	—	1	16	4	17	27	20	33	25	20	11	4	5	183	41	68.9
68.5	1	—	7	11	16	25	31	34	48	21	18	4	3	—	219	49	68.2
67.5	—	3	5	14	15	36	38	28	38	19	11	4	—	—	211	33	67.6
66.5	—	3	3	5	2	17	17	14	13	4	—	—	—	—	78	20	67.2
65.5	1	—	9	5	7	11	11	7	7	5	2	1	—	—	66	12	66.7
64.5	1	1	4	4	1	5	5	—	2	—	—	—	—	—	23	5	65.8
<64.0	1	—	2	4	1	2	2	1	1	—	—	—	—	—	14	1	—
Totals	5	7	32	59	48	117	138	120	167	99	64	41	17	14	928	205	—
Medians	—	—	66.3	67.8	67.9	67.7	67.9	68.3	68.5	69.0	69.0	70.0	—	—	—	—	—

Source: Galton (1886a).

Allein die äußere Form dieser Tabelle suggerierte eine symmetrische Beziehung, in der sich sowohl zeilen- als auch spaltenweise Normalverteilungen zeigten. Nach einer geringfügigen Glättung stellte er folgendes fest: "I then noticed [...] that lines drawn through entries of the same value formed a series of concentric and similar ellipses."¹³⁹ Die Achsen dieser Ellipsen konstanter Häufigkeit gingen durch den gemeinsamen Mittelwert der Daten. Die Punkte, an denen jede Ellipse eine horizontale Tangente schnitt, lagen auf einer geraden Linie mit einer Steigung von 1/3. Die Verbindungslinien aller vertikalen Tangenten lieferte dagegen eine Linie mit einer Steigung von 2/3.¹⁴⁰ Hier lag offensichtlich eine bivariate Normalverteilung vor, und mit den angegebenen Daten konnte Galton nun einen Mathematiker beauftragen, die entsprechende Verteilung herzuleiten.¹⁴¹

Die Folgerung war, daß die Beziehung, die er bei den Erbsen gefunden hatte, keinen besonderen Umstand, der mit dem Gegenstand zusammenhing, darstellten, sondern auf der Tatsache beruhte, daß die Daten bivariat normalverteilt waren. Eine vielleicht noch wichtigere Einsicht war, daß die Beziehung symmetrisch war, also auch nicht notwendigerweise eine kausale Beziehung unterstellt werden mußte. Der nächste Schritt konnte daher darin bestehen, nicht Regressionen, sondern "Ko-Relationen" zu untersuchen. Dazu boten sich Daten über Brüder an.¹⁴²

1888 präsentierte Galton dann der Royal Society einen Vortrag über *Co-relations and their measurement, chiefly from anthropometric data*, in dem er (hier anhand von Vorderarm- und Kopflängen) diesen Sachverhalt deutlich zur Sprache brachte:

¹³⁹ Francis Galton, Regression towards mediocrity in hereditary stature, in: *Journal of the Anthropological Institute* 15 (1886), S. 246-263, zitiert nach Stigler (1986, S. 285).

¹⁴⁰ Diese beiden Linien sind nichts anderes als die Regressionslinien.

¹⁴¹ Galton selbst war nach eigenen Aussagen kein guter Mathematiker.

¹⁴² Zu den Details siehe Stigler (1986, S. 290ff). Interessanterweise verwendete er aus theoretischen und rechentechnischen Gründen hier nicht arithmetische Mittel und quadrierte Abweichungen, sondern Median und Mediane der Abweichungen.

“It is easy to see that co-relation must be the consequence of the variations of the two organs being partly due to common causes.”¹⁴³

Von Galton könnte nun ein direkter Weg zur modernen Statistik führen, doch muß zunächst eine parallele Entwicklung erwähnt werden, die für den weiteren Verlauf von Bedeutung ist. Die Studien Quetelets und Galtons bewegten sich konzeptionell in dem von Bayes, Laplace und Poisson vorgegebenen Rahmen einer inversen Wahrscheinlichkeit, der, wie wir festgestellt hatten, die Statistik bis zum Ende des 19. Jahrhunderts dominierte.¹⁴⁴ Gegen diese Konzeption begann sich seit den 1840er Jahren von philosophischer Seite her Widerspruch zu formieren.¹⁴⁵ 1842 erschien der *Versuch einer Kritik der Principien der Wahrscheinlichkeitsrechnung* von Jakob Friedrich Fries, ein Jahr später John Stuart Mills *System of Logic* und Antoine Augustin Cournots Werk *Exposition de la théorie des chances et des probabilités*, 1844 schließlich Robert Leslie Ellis' *On the foundations of the theory of probabilities*.¹⁴⁶ Ihren Höhepunkt fand diese Kritik in den 1854 und 1866 erschienenen Arbeiten von George Boole, *An investigation of the laws of thought on which are founded the mathematical theories of logic and probabilities* und von John Venn, *The logic of chance*.¹⁴⁷

Um die Jahrhundertwende, spätestens jedoch in den ersten Dekaden unseres Jahrhunderts, gelang dieser häufigkeitstheoretischen Sichtweise der Durchbruch. Lehrbücher der Wahrscheinlichkeitstheorie, die sich an der Laplaceschen Konzeption orientierten, wurden immer seltener und verschwanden schließlich völlig.¹⁴⁸ Wie dies genau vor sich ging, ist bis heute nicht geklärt.¹⁴⁹

¹⁴³ Zitiert nach Stigler (1986, S. 297). Auch diese Werte wurden auf intuitivem, graphischem Wege ermittelt. Galton standardisierte beide Variablen (Körpergröße und Ellenlänge erwachsener Männer), unterteilte die Wertebereiche in Gruppen, ermittelte den Median jeder Zelle und zeichnete diese Werte in ein Streudiagramm. Anschließend legte er durch diese Punkte nach Augenmaß eine Linie. Deren Steigung ($r = 0.8$) bildete den Korrelationskoeffizienten.

¹⁴⁴ Siehe oben.

¹⁴⁵ Siehe dazu Kamlah (1987), Krüger (1987a), Krüger (1987b), Porter (1986, Kap. 3).

¹⁴⁶ Vgl. Porter (1986, S. 77f), Dale (1991, passim), zur Kritik von Fries, Ellis und Cournot siehe Krüger (1987a, S. 68-70, 72f).

¹⁴⁷ Nach Krüger (1987a, S. 70) war Venn der erste, der Ellis' Häufigkeitssicht der Wahrscheinlichkeit zu einem vollständigen philosophischen System ausbaute. Interessanterweise stellte sich Venn gerade gegen die Anwendung wahrscheinlichkeitstheoretischer Aussagen hinsichtlich gesellschaftlicher Phänomene, wie sie von Quetelet und Buckle propagiert worden waren. Die häufigkeitstheoretische Sicht implizierte, daß die betrachteten Ereignisse zum einen homogen sein mußten, zum anderen, daß die einzelnen Gründe für das Zustandekommen eines spezifischen Ereignisses unbekannt waren und eben nur dessen (gleichbleibende) Verteilung auf lange Sicht angegeben werden konnte. In den von Quetelet oder Buckle betrachteten Zusammenhängen seien aber Gründe für das Zustandekommen eines spezifischen Ereignisses (“outcome”) anzugeben. Andererseits änderten sich Zusammenhänge im Laufe der Zeit, so daß eine häufigkeitstheoretische Sicht auch aus diesem Grund ausscheide. Vgl. Porter (1986, S. 175).

¹⁴⁸ Vgl. Kamlah (1987, S. 111). Eines der letzten Bücher in dieser Tradition war E. Czubers erstmals 1902, 1908 in einer zweiten Auflage erschienene *Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichsrechnung, Statistik und Lebensversicherung*. Dagegen war Venns *Logic of Chance* 1876 in der zweiten, 1888 in der dritten Auflage erschienen, und auch andere schlossen sich dieser Interpretation an, wie etwa das 1886 erschienene Lehrbuch *Algebra* von G. Chrystal.

Eine genaue Analyse der möglichen Gründe für diesen konzeptionellen Wandel – die wir hier nicht leisten können – wird erheblich durch die Tatsache erschwert, daß die zentralen Personen, die den Grundstein der modernen Statistik legten, K. Pearson und F. Y. Edgeworth, beide im Laufe ihres Werkes diesen Wandel selbst vollzogen haben und somit zentrale statistische Konzepte zum Teil mit einer inversen, zum Teil mit einer direkten Wahrscheinlichkeitskonzeption verbunden sind.¹⁵⁰ Mit Edgeworth und Pearson vollzog sich auch eine vorläufige Trennung zwischen den bis dahin verwobenen Beziehungen zwischen der Statistik und deren Anwendungsgebieten im Rahmen natur- oder gesellschaftswissenschaftlicher Phänomene,¹⁵¹ bevor die Ökonometrie ein halbes Jahrhundert später die naturwissenschaftliche Entwicklung wieder adaptieren sollte.¹⁵²

c. Beginn der modernen Statistik: F. Y. Edgeworth und K. Pearson

Die ersten Bausteine der “modernen” Statistik wurden von F. Y. Edgeworth gelegt.¹⁵³ Seine Bedeutung liegt vor allem darin, daß er die konzeptionellen und

¹⁴⁹ Die hier zitierten Arbeiten zeichnen sich in diesem Punkt nicht durch große Klarheit aus. Kamlah (1987, S. 113) erwägt externe Gründe als mögliche (Teil-) Erklärung: “The gestalt switch was caused at least partly by external reasons. On the one hand the increasing investigation of probability distributions, i.e., of the ‘a posteriori probabilities’, led to the impression that the ‘a priori probabilities’ are only unimportant special cases of quantities that usually have to be determined by experience.” Seiner Ansicht nach habe die Aufdeckung probabilistischer Gesetzmäßigkeiten in der Physik, wie z. B. die statistischen Gesetze der Brownschen Bewegung oder des radioaktiven Zerfalls, zunehmend zu der Überzeugung geführt, daß Wahrscheinlichkeiten physikalische Größen seien. Dies sind allerdings Gründe, die in bezug auf die Anwendung auf biologische oder sogar soziale Phänomene wohl kaum zutreffen dürften.

¹⁵⁰ Wie wir in Abschnitt d. sehen werden, geben noch zahlreiche Ansätze von R. A. Fisher bis heute Anlaß zu Kontroversen hinsichtlich seiner Einschätzung dieses Problems.

¹⁵¹ Wie schwierig zu diesem Zeitpunkt eine Trennung der einzelnen Wissenschaften von der Statistik ist, belegen unter anderem die Richtungskämpfe, die sich zwischen 1830 und 1890 auf institutioneller Seite abspielten. Eine 1833 von Charles Babbage der *British Association for the Advancement of Science* vorgeschlagene Neugründung einer Sektion “Statistics” wurde zunächst abgelehnt und erst ein Jahr später eingerichtet, als A. Quetelet und R. Malthus sich auf einem Treffen zur Armengesetzgebung äußerten. 1856 wurde diese Sektion in “Statistics and Economics” umbenannt. Die Gründung der Sektion fand keine ungeteilte Zustimmung. Ihr wurde vorgeschrieben, sich nicht mit allgemeinen politischen oder ökonomischen Fragen zu befassen. Zunächst wurde die Präsidentschaft selten von Ökonomen ausgeübt. Dies änderte sich erst, als Galton 1877 versuchte, die Sektion aus der Gesellschaft zu lösen. Die Verflechtung mit der Statistik rief bei Ökonomen zunehmend Unwillen hervor, und so wurde 1890 schließlich die *British Economic Association* gegründet. Die Statistiker empfanden diese Situation bereits seit etwa 1833 als unbefriedigend und gründeten im folgenden Jahr die *London Statistical Society*. Seit 1837 gab die Gesellschaft ihre eigene Zeitschrift heraus, in der in der Regel ökonomisch-statistische Problemstellungen diskutiert wurden. Vgl. Koot (1987, S. 61-63). Der Präsident des Internationalen Statistischen Kongresses 1860 war Prince Consort, ein Schüler Quetelets. Vgl. Porter (1986, S. 165).

¹⁵² Siehe dazu unten, Abschnitt B 3 b.

¹⁵³ Francis Ysidro Edgeworth [1845-1926], Sohn eines irischen *landlord* und entfernter Verwandter von F. Galton, studierte seit 1862 am Trinity College in Dublin, seit 1867 in Oxford, klassische Literatur, anschließend Handelsrecht sowie (im Selbststudium) Mathematik. 1880 wurde er *lecturer* für Logik am *King’s College* in London. Edgeworth ist heute vor allem aufgrund seiner wirtschaftstheoretischen Arbeiten, in denen er Lehren der klassischen

intuitiven Ansätze Galtons aufgegriffen, formalisiert und von ihrer scheinbaren Gebundenheit an eugenische Probleme gelöst hat. Edgeworths Beitrag zur Entwicklung der Statistik begann 1883/1884 mit einer Reihe von Arbeiten über *“The law of error”*, *“The method of least squares”*, *“The physical basis of probability”*, *“The philosophy of chance”*, *“On the reduction of observations”*, *“A priori probabilities”* und mündete 1885 in vier zentrale Arbeiten über *“Observations and statistics: An essay on the theory of errors of observation and the first principles of statistics”*, *“Methods of statistics”*, *“On methods of ascertaining variations in the rate of births, deaths, and marriages”* und *“Progressive means”*.¹⁵⁴

Wie Galton ging auch Edgeworth davon aus, daß sich durch geeignete Einteilungen stabile Verteilungen in nahezu allen Bereichen finden lassen (und somit auch Signifikanztests angebracht seien). Er propagierte den Nutzen statistischer Ansätze für eine Vielzahl unterschiedlichster Anwendungsgebiete: in bezug auf Heiraten, Sterbefälle, diverse Wirtschaftsstatistiken, Clubbesuche, dem Versmaß in Virgils Gedichten, bis hin zur Untersuchung von An- und Abflügen auf und von Wespenestern.¹⁵⁵ In all diesen Beispielen verwendete er Techniken, die nicht neu waren. Neu dagegen war der konzeptionelle Rahmen, daß diese Techniken auf solche Probleme Anwendung finden konnten. Dabei berief er sich explizit auf Galton; auch dessen Beispiele (des Galtonschen Brettes) – und selbst das Brett, das er sich von Galton zu Demonstrationszwecken auslieh – verwendete er. Im Gegensatz zu den klassischen Arbeiten der Astronomen und Geodäten setzte er den Weg Quetelets fort. Die interessierenden statistischen Größen waren selbst Subjekt von Verteilungen, wie etwa im Beispiel des Galtonschen Brettes oder der durchschnittlichen Größe einer Frucht an einem zufälligen Ort im Garten, nicht deren Meßfehler:

“The classical theory provided a rationale for the appearance of probability distributions (in particular, normal curves), where the goal was determining fixed, objectively defined quantity from observations that were subject to measurement errors. Now these distributions could be rationalized for social data at arbitrary levels of classification, just as a quincunx could be cut at arbitrary levels. And the distributions, once rationalized, themselves served as objects for comparison. The conception was not yet fully realized in 1885, but it was a major advance from Quetelet, for whom measurement error analogy had been necessary.”¹⁵⁶

Edgeworth legte Wert darauf, *Beobachtungen* (observations) und *Statistiken* (statistics) zu unterscheiden. In der Astronomie und Geodäsie, in dessen Rahmen die statistischen Konzepte entwickelt wurden, bestanden Variationen lediglich aus Meßfehlern, die man mit gutem Grund als geringfügige und zufällige Überlagerungen von Konstanten ansehen konnte.¹⁵⁷ Im Rahmen der sozioökonomischen

Nationalökonomie mit Ansätzen der Grenznutzenschule verband, bekannt (*“Edgeworth-Diagramm”*).

¹⁵⁴ Bibliographische Angaben zu den einzelnen Titeln finden sich bei Stigler (1986, S. 379ff).

¹⁵⁵ Hierzu führte er auf dem Landsitz seiner Familie Untersuchungen durch. Vgl. Porter (1986, S. 266).

¹⁵⁶ Stigler (1986, S. 313).

¹⁵⁷ Es bleibt offen, wie in diesem Zusammenhang die Daten Galtons einzuordnen sind. Hier

Daten standen keine Messungen im Mittelpunkt, sondern die Daten selbst (die Statistiken) und ihre Variation waren von Interesse. Edgeworth verdeutlichte dies anhand der Bestimmung eines Mittelwertes. Dazu fragte er nach der Bedeutung des Wahrscheinlichkeitsbegriffes: Was war in diesem Zusammenhang damit gemeint? Seiner Ansicht nach konnte diese Frage nur im Sinne von Bayes und Laplace beantwortet werden. Wie Laplace vertrat er eine inverse Wahrscheinlichkeitskonzeption, und wie Laplace bevorzugte er in bezug auf Signifikanztests jedoch eine frequentistische Sicht.¹⁵⁸ Einzige Voraussetzung hierfür war, daß die zu untersuchende Größe einem konstanten Gesetz unterlag. Der beste Wert für die Wahl eines Mittelwertes sei derjenige, der die *A-posteriori*-Wahrscheinlichkeit maximierte. Für einen Test der Differenz zweier Mittelwerte empfahl Edgeworth 1885 folgende Vorgehensweise:

“In order to determine whether the difference between two proposed Means is or is not accidental, form the probability curve under which the said difference, supposing it were accidental, would range. Consider whether the difference between the observed Means exceeds two or three times the modulus of that curve. If it does, the difference is not accidental.”¹⁵⁹

Auch diesen Test wandte Edgeworth auf eine Reihe von Daten aus den verschiedensten Bereichen an (Anthropometrie, der “Moral”statistik, ökonomische Daten¹⁶⁰ etc.).

Der Regressions- und Korrelationsrechnung Galtons stand Edgeworth zunächst überraschenderweise skeptisch gegenüber. In einer Besprechung von *Natural Inheritance in Nature* bezweifelte er, ob sich die dort vorgeschlagenen Verfahren auf breiter Ebene durchsetzen würden, ob es viele Daten gäbe, die sich auf diese

steht zwar auch die Variation im Vordergrund und ein fester Wert wie die Position eines Planeten existiert nicht physikalisch (etwa bei Erbsen); die Annahme eines solchen festen “Parameters” ist aber theoretisch eher zu rechtfertigen als im sozialwissenschaftlichen Kontext. Quetelet ging noch von solch einer Analogie aus.

¹⁵⁸ “He gave careful consideration to the objective frequentist or sampling theory view (he called it ordinary induction, assumed inversion, or the indirect method), but in the end he adopted the inverse probability, or Bayesian, view after giving cogent refutations to criticisms of this view by Cournot, Boole, and Venn. [...] Like Laplace, Edgeworth embraced inverse probability for neither metaphysical (his priors were grounded on experience) nor dogmatic reasons. Like Laplace, Edgeworth was inconsistent in his application of inverse probability, reverting to sampling distributions when comparing means by significance tests”. Stigler (1986, S. 310). Porter (1986, S. 265) ist der Ansicht, daß Edgeworth in diesen Arbeiten besonders an der Beziehung zwischen objektiven Häufigkeiten und subjektiven Überzeugungen interessiert war, führt dies aber nicht aus. Zu den mathematischen Details in verschiedenen speziellen Zusammenhängen vgl. Dale (1991, S. 342-348). Nach Edgeworths Ansicht war eine A-priori-Gleichverteilung in vielen Situationen aufgrund von Erfahrung berechtigt.

¹⁵⁹ Zitiert nach Baird (1988, S. 468). Dieser Test entspricht einem zweiseitigen Test mit einem Signifikanzniveau von 0.005. Nach Stigler (1986, S. 312) wurden solche Tests auch von anderen Forschern, auch in bezug auf soziale Daten, angewandt, doch blieb dies die Ausnahme und erreichte nicht die abwägende Qualität der Arbeiten Edgeworths.

¹⁶⁰ Unter anderem auch mit Daten aus Jevons *Investigations in Currency and Finance*. Das Buch hatte Edgeworth zuvor rezensiert. Im selben Jahr untersuchte Edgeworth die Frage, ob ein Trend der englischen Sterberaten von 1838 bis 1878 (berechnet als Kleinstquadratgerade) eine signifikante Steigung aufwies oder nicht.

Art und Weise behandeln ließen. Wenige Jahre später änderte er jedoch seine Meinung.¹⁶¹ Er nahm anhand der Daten, die Galton für die intuitive Bestimmung des Korrelationskoeffizienten verwendet hatte, einige neue Berechnungen vor und entwickelte in einem theoretischen Rahmen dessen Ideen weiter. Eigene Anwendungen lieferte er jedoch zunächst nicht.

Den Grundstein der modernen Statistik legte schließlich K. Pearson,¹⁶² indem er Galtons und Edgeworths Ansätze systematisierte. Auf K. Pearson gehen neben technischen Einzelheiten vor allem die folgenden Errungenschaften zurück:¹⁶³

- die Anwendung statistischer Verfahren auf einen breiten Problemkreis,
- die Ableitung von partiellen und multiplen Korrelationskoeffizienten,¹⁶⁴
- die Definition der gemeinsamen Verteilung von Schätzungen,
- der χ^2 -Anpassungstest,¹⁶⁵
- die auf diesen Test aufbauende Analyse von Kontingenztafeln,
- ein System von Häufigkeitskurven und
- die Methode der Momente als Schätzsystem.

Sein Interesse für die mathematische Statistik wurde nach Stigler nicht durch Galton, sondern durch Edgeworth geweckt. Pearson hielt zwar bereits 1889 Abendkurse über Galtons *Natural inheritance*, war aber zunächst skeptisch; in seiner 1892 erschienenen *Grammar of science* fand Galton keine Erwähnung.¹⁶⁶ Sein erster eigener Beitrag zur Statistik, der seinen späteren Ruhm begründete, bedeutete zunächst einen Rückschritt gegenüber Galton und Edgeworth: sein

¹⁶¹ Der Grund für diesen Meinungswechsel ist nach Stigler (1986, S. 316) unklar. Die bi- und trivariate Normalverteilung war als mathematische Größe bereits von Laplace abgeleitet und 1846 von A. Bravais näher untersucht worden. Hinsichtlich ihrer Anwendung beschränkten sich die wenigen Untersuchungen auf Probleme der Ballistik.

¹⁶² Karl Pearson [1857-1963] studierte bis 1879 Mathematik in Cambridge, anschließend ein Jahr in Heidelberg und Berlin Physik, Metaphysik, Darwinismus, Mittelhochdeutsch und Folklore. Daraufhin ging er zurück nach London, studierte dort Rechtswissenschaft und wurde 1881 als Anwalt zugelassen, übte den Beruf aber nicht aus. 1884 konnte er für eine "Goldsmid Professorship of Engineering and Applied Mathematics" am *University College* London gewonnen werden. Zu diesem Zeitpunkt war Pearson bereits als Autor einer Essaysammlung, die sich hauptsächlich mit sozialen Fragen beschäftigte, bekannt sowie durch sein Buch *Grammar of Science*. Er war Mitglied des "Men's and Woman's Club", der sich für die Gleichstellung der Frauen einsetzte und bezeichnete sich selbst als Sozialist. Gleichzeitig bekannte er sich zu einem (zur Jahrhundertwende nicht selten anzutreffenden) Sozialdarwinismus, der durchaus rassistische Elemente aufwies. Die Statistik wurde vergleichsweise spät (um 1894) zu seinem Hauptinteresse.

¹⁶³ Vgl. Lancaster (1988, S. 707).

¹⁶⁴ "Francis Galton invented correlation, but Karl Pearson was chiefly responsible for its development and promotion as a scientific concept of universal significance." Aldrich (1995, S. 364).

¹⁶⁵ Das Problem der Freiheitsgrade wurde jedoch erst von R. A. Fisher geklärt. Siehe dazu Bartlett (1965, S. 396).

¹⁶⁶ Vgl. Stigler (1986, S. 300-305).

System von Häufigkeitsverteilungen, die er in zwei umfangreichen Beiträgen in den *Philosophical Transactions of the Royal Society* unter dem Titel "*Contributions to the mathematical theory of evolution*" veröffentlichte und die zu seiner Aufnahme in die Gesellschaft führten.

Die Frage nach der Form von Häufigkeitsverteilungen war seit dem Ende des 18. Jahrhunderts ein grundlegendes Thema. Es herrschte die allgemeine Überzeugung vor, daß einzelne Phänomene, die homogen im Sinne von vielen, individuell unbedeutenden Einflußfaktoren waren, einer Normalverteilung folgen mußten. Dennoch wurde die Normalverteilung nicht von jedermann als universal gültige Verteilung angesehen. Im Laufe der Jahre häuften sich Datensammlungen, die eine Reihe von "schiefen" Verteilungen implizierten. Diese Tatsache wurde von Edgeworth, insbesondere aber von Pearson als Herausforderung angesehen, und letzterer entwickelte schließlich eine "Kurvenfamilie" mit jeweils vier Parametern, anhand derer Daten über ihre ersten vier Momente verschiedenen Typen von Kurven zugeordnet werden konnten.

Pearson lieferte nicht nur die Formeln, sondern auch eine Fülle von Anwendungsbeispielen (die Verteilung des Luftdrucks, die Größe von Schulkindern bzw. von Krebsen, Statistiken über den Pauperismus, über Ehescheidungen etc.) und zeigte, daß sich diese Größen durch sein System in hoher Übereinstimmung darstellen ließen. Damit ging er sogar über Quetelets Ansatz hinaus. Nicht nur die Daten, die eine Normalverteilung aufwiesen, folgten einem einheitlichen Verteilungsgesetz, ohne daß Gruppen / große Faktoren isoliert werden mußten, sondern auch noch zahlreiche andere, deren Verteilung zwar schief, aber deswegen nicht weniger gesetzmäßig war. Sollte dies zutreffen, war die Suche nach verursachenden Faktoren, wie sie von Galton eingeführt wurde, hinfällig:

"The law of frequency is based on the assumption of perfect ignorance of causes, but we rarely are perfectly ignorant, & where we have any knowledge it ought of course to be taken into account."¹⁶⁷

Auch Pearsons Anwendung auf Bereiche, die nicht unbedingt ein konstantes Verteilungsgesetz nahelegten, wurde kritisiert.¹⁶⁸

"[...] I see that there are many cases of 'skew' variation: but all cases which he has given, of variation with an unmistakably skew frequency, are taken from phenomena which are changing with a rapidity much greater than that of any organs in crabs, or such creatures. Pauperism, divorces, and the like, have only been invented, in their present form, for a short time, and as he himself shows, the maximum frequency changes its position at least in ten years."¹⁶⁹

Und schließlich brachte Edgeworth das wohl wichtigste Gegenargument: Den zahlreichen Formen, die sich durch Pearsons Häufigkeitskurven anpassen ließen,

¹⁶⁷ F. Galton in einem Brief an K. Pearson vom 18.11.1893, zitiert nach Stigler (1986, S. 336). Hervorhebung im Original.

¹⁶⁸ Trotz der Kritik gehörten Pearsons Häufigkeitskurven bald zum Standardrepertoire der Statistik.

¹⁶⁹ W. F. R. Weldon in einem Brief an F. Galton vom 27.1.1895, zitiert nach Stigler (1986, S. 337).

fehlte eine theoretische Begründung; es waren rein empirische Konstrukte. Wenn sich eine Häufigkeitsverteilung nicht durch eine Normalverteilung repräsentieren ließ, so konnte das Konzept einer Verursachung durch eine große Zahl zufälliger Ursachen nicht greifen. Genau dieser letzte Punkt war jedoch nach Stigler (1986, S. 339) auch nicht die Intention Pearsons. Dieser habe nach C. S. Peirce eine Wissenschaftsphilosophie vertreten, die sich am Kantschen Nominalismus orientiert habe. Demnach waren die Häufigkeitskurven für Pearson nur gedankliche Konstrukte, die empirische Evidenz zusammenfassen, jedoch keine Aussagen über mögliche Ursachen liefern sollten. Dennoch suchte Pearson auch hier nach einem formalen Kriterium zur Beurteilung der Abweichung der empirischen Verteilungen von seinen Häufigkeitskurven und fand es schließlich in dem von ihm 1900 vorgestellten χ^2 -Test.

Einen weiteren Beitrag zur modernen Statistik lieferte Pearson auf dem Gebiet der Korrelation. Er betrachtete zwei Größen, die bivariat normalverteilt waren, leitete daraus den Korrelationskoeffizienten sowie dessen A-posteriori-Verteilung¹⁷⁰ ab (in Abhängigkeit von den empirischen Standardabweichungen) und systematisierte die bisherigen Erkenntnisse. Der theoretischen Ableitung folgten eine Reihe von Anwendungsillustrationen, die er von Galton übernahm. Eine Synthese kann man schließlich in einer umfangreichen, 1898 zusammen mit N. G. Filon verfaßten Arbeit sehen.¹⁷¹

Einer Anwendung auf soziale Phänomene räumte er keine großen Möglichkeiten ein:

“Personally I ought to say that there is, in my own opinion, considerable danger in allying the methods of exact science to problems in descriptive science, whether they be problems of heredity or of political economy; the grace and logical accuracy of the mathematical process are apt to fascinate the descriptive scientist that he seeks for sociological hypotheses which fit his mathematical reasoning and this without first ascertaining whether the basis of his hypotheses is as broad as that human life to which the theory is to be applied.”¹⁷²

¹⁷⁰ Explizit lehnte K. Pearson die Konzeption einer inversen Wahrscheinlichkeit ab, doch E. S. Pearson vertrat die Ansicht, daß er zumindest in einem Fall implizit diesem Ansatz folgte. Vgl. Pearson (1898). “The basic of the approach used here is a little obscure and there seems to be implicit in it the classical concept of inverse probability.” Pearson (1967, S. 347), zitiert nach Dale (1991, S. 379). Am ausführlichsten hat sich Pearson in seinem Aufsatz *The fundamental problem of practical statistics* zu diesem Problem geäußert und damit bis heute unterschiedliche Interpretationen hervorgerufen. Während Fisher (1922, S. 311) z. B. darin einen Beweis des Bayes-Theorems zu erkennen glaubte, hält dies Dale (1991, S. 388) für eine “totally inaccurate observation”. Für weitere Interpretationen vgl. ebda., S. 377-391. Nach Stigler (1986, S. 345) arbeitete Pearson mehrfach “[...] (implicitly) in a Bayesian framework”.

¹⁷¹ Vgl. Pearson/Filon (1898). Auch hier finden sich noch zwei verschiedene Schätzprinzipien: der (bayesianisch aufgefaßte) Maximum-Likelihood-Ansatz für den Korrelationskoeffizienten, wobei die Standardabweichungen durch die entsprechenden Momente geschätzt werden sollten und die ausschließliche Verwendung von Momentenschätzern für die Parameter der Häufigkeitskurven.

¹⁷² Pearson (1898, S. 1f), zitiert nach Stigler (1986, S. 304).

Diese Übertragung wurde schließlich von Pearsons Schüler G. U. Yule geleistet.¹⁷³ Yule wandte zunächst die von Pearson entwickelten Häufigkeitskurven auf die Anzahl der Armen in 580 Verwaltungseinheiten in England und Wales 1891 an, um diese zu typisieren.¹⁷⁴

Darüber hinaus untersuchte er eine Zusammenhangstruktur. Die Frage war, inwieweit der Anteil der Armen in einem Distrikt mit der Fürsorgestruktur zusammenhing.¹⁷⁵ Yule fand hier einen "signifikanten" Zusammenhang,¹⁷⁶ den er jedoch als "suggestive" bezeichnete, da die Verteilungen der beiden Größen eine deutliche Schiefe aufwiesen.¹⁷⁷ In einem nächsten Schritt bestimmte er zwischen diesen beiden Größen eine "Regressionslinie", indem er die Abstände zwischen dieser Gerade und den Daten minimierte.¹⁷⁸ Er erkannte, daß dieser Ansatz leicht auf höhere Dimensionen erweiterbar war und zu eben jenem "Normal-" Gleichungssystem führte, das Gauss mehrere Jahrzehnte vorher eingeführt hatte.¹⁷⁹ Von diesem Punkt aus war es nur noch ein technischer Schritt, den Ansatz auf mehr als zwei Variablen auszudehnen.¹⁸⁰

¹⁷³ George Udny Yule [1871-1951] entstammte einer schottischen Gelehrtenfamilie, studierte zunächst Ingenieurwesen in London und widmete sich anschließend ein Jahr in Bonn dem Studium elektrischer Wellen bei Heinrich Hertz. Aus dieser Zeit stammte seine Vorliebe für graphische Darstellungen (vgl. die Einleitung von Kendall/Stuart (1971, S. 1)). 1893 kehrte er nach London zurück und bekam dort von K. Pearson, den Yule als Student kennengelernt hatte, eine *demonstratorship* am *University College* (London) angeboten; anschließend wurde er dort *assistant professor*. 1899 verließ er das College und ging an das *Department of the City and Guilds des London Institute*. Zwischen 1902 und 1909 hielt er, wiederum am *University College*, die *Newmarch Lectureship in Statistics* (an der Fakultät für Agrarwissenschaften). 1912 wurde er *lecturer* für Statistik an der Universität Cambridge, wo er, abgesehen von Tätigkeiten für das *War Office* und das Ernährungsministerium zwischen 1915 und 1919, bis 1940 lehrte.

¹⁷⁴ Vgl. Yule (1896a).

¹⁷⁵ Vgl. Yule (1895), Yule (1896b). Mit Fürsorgestruktur ist hier das Verhältnis von unmittelbarer und in Form von Kost und Logis in Armenhäusern gewährter Unterstützung gemeint.

¹⁷⁶ Der Anlaß für diese Untersuchung war eine von C. Booth geäußerte These, daß zwischen diesen beiden Größen kein Zusammenhang bestehe. Yule berechnete zwischen dem Verhältnis der "Out-paupers" zu den "In-paupers" und dem Prozentsatz der Bevölkerung, der Armenfürsorge erhielt, einen Korrelationskoeffizienten von 0.388 mit einer Standardabweichung ("probable error") von 0.022. Dies war vermutlich das erste Mal, daß eine Hypothese in den Sozialwissenschaften aufgrund eines "signifikanten" empirischen Befundes abgelehnt wurde.

¹⁷⁷ In einer folgenden Arbeit wies er nach, daß im Falle einer linearen Beziehung zweier Variablen der Steigungskoeffizient der Regressionsgeraden durch die Größen $p(\sigma_1/\sigma_2)$ bzw. ihre Momente r , s_1 und s_2 geschätzt werden könne. Pearson hatte lediglich den umgekehrten Fall untersucht.

¹⁷⁸ Yule (1897, S. 814) betonte, daß die Verwendung des Kleinstquadratkriteriums nicht aus logischen Gründen erfolgte, sondern "solely for convenience of analysis". Auch die Form der Beziehung ergäbe sich nicht theoretisch, sondern: "It must be understood that we only take a *linear* characteristic relation because it is the simplest possible form to calculate." (S. 816). Hervorhebung im Original.

¹⁷⁹ "[...] so named by Gauss, as if to predict that Yule would reconcile them with 'normal correlation' nearly ninety years later!" Stigler (1986, S. 350f). Ein erheblicher Vorteil bestand auch darin, daß durch die Anlehnung an die im Rahmen der Astronomie und Geodäsie entwickelten Ansätze eine Reihe von Algorithmen übernommen werden konnten. Dort bestand bereits ein voll ausgebautes Instrumentarium (s. o.).

Pearson lehnte diesen Weg ab. Die Idee, daß der Zusammenhang zwischen zwei Variablen nicht nur durch eine Korrelation, sondern durch eine lineare Funktion gegeben sein könnte, wollte Pearson allenfalls für die Physik, nicht aber für die Biologie und schon gar nicht für die Sozialwissenschaften gelten lassen:

“In physics you know by experience that the finer your methods of observation and your powers of observation, the more nearly you get your two variables related by a single valued equation and you are justified in trying to find the value of its constants [...]. The key to your method is, such a relation between the two variables actually exists in nature, it is the axiom from which you start. In biology you start with the exact opposite – no such single valued relation exists, but I understand by correlation the theory which endeavours to supply its place. [...] In biology there might possibly be a single-valued function relation, if we could take account of thousands of variables, but we never attempt in our observations to make any approach to such a relation.”¹⁸¹

Die richtige Vorgehensweise habe seiner Ansicht nach darin zu bestehen, “frequency surfaces” zu berechnen und keine einfache Relation. Eine solche sei lediglich anzunehmen, wenn die bivariate Häufigkeitsverteilung einer Normalverteilung entspräche.¹⁸²

Wir wollen zum Abschluß dieses Abschnittes noch einige Hinweise auf die unterschiedlichen Intentionen des Korrelationskonzeptes von Pearson und Yule geben, die vor kurzem von Aldrich näher untersucht worden sind.¹⁸³ Pearson und Yule interpretierten den Begriff der Korrelation unterschiedlich. Pearson hatte seinen Ansatz der Korrelation mit dem Begriff der Kausalität verbunden und in seiner *Grammar of Science* (in der dritten Auflage 1910) den Begriff der Kausalität durch den der Korrelation ersetzen wollen:

“It is this conception of correlation between two occurrences embracing all relationships from absolute independence to complete dependence, which is the wider category by which we have to replace the old idea of causation.”¹⁸⁴

Lediglich eine Ausnahme sei hier zu beachten: der Fall der “spurious correlation”. Pearson untersuchte in zunehmenden Maße Daten, die seiner Ansicht nach die fundamentale Bedeutung der Korrelation bestätigten, fand aber auch immer häufiger Situationen vor, in denen eine Korrelationsanalyse in die Irre führte. Er

¹⁸⁰ Vgl. Yule (1899).

¹⁸¹ In einem Brief an Yule, September 1896, zitiert nach Stigler (1986, S. 351). In späteren Arbeiten hat er seine ablehnende Haltung relativiert.

¹⁸² Stigler (1986, S. 353) begründet Pearsons Ablehnung mit dessen Ausrichtung auf Fragestellungen, in denen es sinnvoller sei, Korrelationen statt Regressionen zu untersuchen, wie z. B. die Beziehung zwischen Kopf- und Fußgrößen, für die eine kausale Fragestellung nicht angemessen war. Ob dies als Erklärung ausreicht, ist allerdings zu bezweifeln.

¹⁸³ Aldrich (1995).

¹⁸⁴ Karl Pearson, *The Grammar of Science*, 3. A., Edinburgh 1910, zitiert nach Aldrich (1995, S. 365). An früherer Stelle führte er aus, daß Korrelation in Kausalität übergehe, wenn der Koeffizient den Wert 1 erreiche. Allerdings sei nicht davon auszugehen, daß eine solche in der Realität existiere. Vgl. ebda.

unterschied daraufhin zwischen “spurious correlations” und “organic relationsips”, wobei nur letztere kausal aufzufassen waren.¹⁸⁵

Prinzipiell verstand Pearson unter einer organischen Beziehung (die sich in einer Korrelation ausdrückte) eine Beeinflussung durch gemeinsame Gründe (common causes). Wie Aldrich zu Recht betont, lag eine solche Interpretation im Gegenstand des von Pearson betrachteten Untersuchungsfeldes begründet: dies waren vor allem biologische und anthropometrische Größen. In diesem Zusammenhang lag es in der Tat näher, z. B. von einer gemeinsamen Beeinflussung der Arm- und Beinlänge zu sprechen, als von einer kausalen Abhängigkeit dieser beiden Größen. In solchen Fällen seien die Korrelationen durch eine große Anzahl unabhängiger gemeinsamer Faktoren hervorgerufen. Eine *Scheinkorrelation* könne künstlich hervorgerufen werden, wenn zwei unabhängige Größen miteinander korreliert werden, die nicht in ihrer originalen Metrik, sondern jeweils als Anteil an einer dritten, ebenfalls von beiden unabhängigen Variablen zugrundegelegt werden.¹⁸⁶ Solche “Daten” sah er insbesondere in den Bereichen der Sozial- oder Medizinstatistik gegeben.¹⁸⁷ Weiterhin hatte Pearson festgestellt, daß eine Scheinkorrelation auftreten könne, wenn die Daten aus zwei verschiedenen Populationen stammen, die jeweils unkorreliert sind.

Yule hatte in seinen Korrelationsstudien intensiven Gebrauch von Anteilen gemacht und damit nach Pearson Scheinkorrelationen errechnet. Yule wehrte sich gegen diesen Vorwurf. Seiner Ansicht nach waren bei der Korrelation von Anteilen drei Situationen zu unterscheiden: (1) wenn die Gründe (causes) unmittelbar die absoluten Größen beeinflussen, (2) wenn sie die Anteile beeinflussen und (3) wenn “we have no direct knowledge of the mode of operation of the causes”.¹⁸⁸ In den ersten beiden Fällen seien die Größen zu korrelieren, die gemeinsamen Faktoren unterlagen, also entweder die absoluten oder die relativen Größen. Das bedeutete – im Gegensatz zu Pearsons Ansicht –, daß auch eine Korrelation absoluter Größen irreführend sein konnte.¹⁸⁹ Die Frage, die sich also stellte, war: Wie konnte man davon ausgehen, daß eine errechnete (hohe) Korrelation “echt” war und nicht nur auf einer falschen Skalierung beruhte?

Yules von Pearson abweichende Interpretation der Korrelation entwickelte sich in den Studien über die Armengesetzgebung. In diesem sozioökonomischen Zusam-

¹⁸⁵ Wie Aldrich an anderer Stelle darlegte, war dieses Konzept Pearsons äquivalent zu den von anderen postulierten “Naturgesetzen”. Zur Beziehung der Pearsonschen zu anderen (modernen) Kausalitätskonzeptionen siehe ebda., S. 366.

¹⁸⁶ Pearson hatte in einem Beitrag nachgewiesen, daß im Falle dreier unabhängiger Variablen x_1, x_2, x_3 die *theoretische* Korrelation zwischen x_1/x_3 und x_2/x_3 zwischen 0.4 und 0.5 lag.

¹⁸⁷ Auch in diesem Bereich hat Pearson seinen Standpunkt im Laufe der Zeit geändert: Während er in früheren Arbeiten davon ausging, daß eine “spurious correlation” vorrangig durch den datentransformierenden Forscher verursacht wurde, sah er später (aufgrund größerer Erfahrung) den Fehler des Statistikers nicht mehr darin, künstlich Scheinkorrelation zu verursachen, sondern die vorhandene Scheinkorrelation nicht angemessen beseitigt zu haben.

¹⁸⁸ Zitiert nach Aldrich (1995, S. 367).

¹⁸⁹ Für diesen Fall entwickelte Yule eine analoge Formel zu derjenigen Pearsons, nun aber für absolute Größen, wenn Anteile unkorreliert sind.

menhang bestand das Interesse nicht in der Suche nach gemeinsamer Variation, sondern nach Ursachen und Wirkungen:

“This statement does not say either that the low mean proportion of out-relief is the cause of the lesser mean pauperism or vice versa: such terms seem best avoided where one is not dealing with a catena of causation at all. To use a simile, due I believe to Professor Marshall [...] the case is like that of a lot of balls – say half a dozen – resting in a bowl. Then you cannot say that the position of ball No. 3 is the cause of the position of No. 5 nor the reverse. But the position of 3 is a function of the positions of all the others including 5; and the position of 5 is a function of the positions of all the others including 3: hence variations in the positions of the two balls will be correlated, and it is to this term that I prefer to adhere. To be quite clear, I do not mean simply that out-relief determines pauperism in one union and pauperism out-relief in another, so that you cannot say which is which in the average: but I mean that out-relief and pauperism mutually react in one and the same union.”¹⁹⁰

Kurz darauf ging Yule einen Schritt weiter und untersuchte die Möglichkeit, daß administrative Handlungen (Wahl der out-door-relief-Anteile) den Armenanteil beeinflussen konnten. Er untersuchte nun auch *partielle* Korrelationen. Wenn eine Beziehung zwischen zwei Variablen auch dann noch bestand, wenn eine dritte, diese beiden vermutlich beeinflussende Variable konstant gehalten wurde, konnte man eine Scheinkorrelation ausschließen.

Die generelle Frage bei all diesen Überlegungen lautete: Bezog sich die Inferenz auf eine Population oder auf Gesetze? Im Falle Pearsons war dies, wie im übrigen auch später bei Fisher, klar. Bei der Untersuchung biologischer Daten war das Ziel die Eruierung von Gesetzmäßigkeiten. Schwieriger gestaltete sich die Situation bei den von Yule untersuchten sozioökonomischen Daten oder Untersuchungen, wie diejenige Gossets über Korrelationen zwischen Krebsraten und Apfelkonsum; hier war zumindest eine explorative Komponente enthalten.¹⁹¹ Die Interpretation eines Korrelationskoeffizienten konnte nach Yule nur hypothetisch sein, da in der Regel eine Vielzahl alternativer Erklärungen möglich waren, deren Unterscheidung die Statistik nicht leisten könne. Dieses Problem sollte sich für inferenzstatistische Interpretationen in der Sozialwissenschaft als fundamental erweisen.

d. Grundlegung der modernen Schätztheorie: R. A. Fisher

K. Pearson hatte die theoretische Statistik in mancher Hinsicht weitergebracht; ihre systematische Fundierung wurde jedoch nach allgemeiner Auffassung erst von R. A. Fisher gelegt.¹⁹² Neben zahlreichen technischen Detaillösungen verdanken wir Fisher in bezug auf die Entwicklung statistischer Konzepte vor allem:¹⁹³

¹⁹⁰ Yule (1895, S. 605). Vgl. auch Aldrich (1995, S. 368).

¹⁹¹ Vgl. ebda., S. 373.

¹⁹² Sir [seit 1952] Ronald Aylmer Fisher [1890-1962] ist zweifellos der bedeutendste Statistiker des 20. Jahrhunderts und hat auch auf dem Gebiet der Genetik Weltruhm erlangt. J. O. Irvin (1963, S. 159) gibt in einem Nachruf die wohl nahezu einhellige Auffassung der statistischen Profession wieder: “Sir Ronald Fisher was the most famous statistician and mathematical biologist in the world.” Fisher entwickelte die Statistik vor allem im Hinblick auf Anwendungen in der Biologie. Er graduierte 1912 in Cambridge in Mathematik, beschäftigte sich

- (1) die Entwicklung der Theorie kleiner Stichproben und damit Stichproben jeder Größe,
- (2) die systematische Entwicklung von Signifikanztests, unter anderem durch die "Entdeckung" der F -Verteilung¹⁹⁴ und die Angabe der "richtigen" Anzahl der Freiheitsgrade bei der Verwendung des χ^2 -Tests,¹⁹⁵
- (3) die Entwicklung von Kriterien für eine "gute" Schätzung: neben Erwartungstreue Effizienz, Konsistenz und Suffizienz,
- (4) den Hinweis auf die fundamentale Bedeutung der Likelihoodfunktion für die Inferenz,¹⁹⁶
- (5) eine allgemeine Methode, die Schätzer liefert, die (meistens) automatisch die unter (3) geforderten Bedingungen erfüllt: die Maximum-Likelihood-Methode und die
- (6) Angabe der Qualität eines Schätzers durch Intervalle: die Fiduzialinferenz.

Die ersten beiden Beiträge Fishers zur Statistik waren eine Darstellung der Maximum-Likelihood-Methode für Häufigkeitskurven (1912) und die Ableitung der Stichprobenverteilung des Korrelationskoeffizienten im Jahre 1915. Die folgenden Arbeiten lassen sich in drei Kategorien einteilen:¹⁹⁷ Untersuchungen über die

dort anschließend mit statistischer Mechanik, der Quanten- und Fehlertheorie. Zwischen 1913 und 1915 arbeitete er zeitweise für eine Versicherungsgesellschaft sowie auf einer Farm in Kanada. Aufgrund seiner schlechten Sehstärke (die ihm zeitweise weder Lesen noch Schreiben erlaubte) wurde er nicht zum Militärdienst eingezogen und lehrte zwischen 1915 und 1919 Mathematik und Physik an öffentlichen Schulen. Ab 1919 arbeitete er auf der *Rothamsted Experimental Station* (als einziger Statistiker, nachdem er ein Assistenten-Angebot K. Pearsons abgelehnt hatte), war von 1933 bis 1943 als Nachfolger Karl Pearsons *Galton professor of eugenics* am *University College* in London, anschließend bis zu seiner Emeritierung 1957 Professor für Eugenik in Cambridge. 1959 ließ er sich in Australien nieder.

¹⁹³ Konzeptionelle Überlegungen finden sich vor allem in Fisher (1922), Fisher (1932), Fisher (1935a) und Fisher (1959). Als grundlegender Beitrag zur Entwicklung der modernen theoretischen Statistik wird zumeist Fisher (1925) gewürdigt, obwohl bereits in Fisher (1922) ein Großteil der dort geäußerten Gedanken enthalten ist, wie Geisser (1992) betont. Seine Sicht der Inferenz zusammen mit abschließenden Angriffen auf alternative Konzeptionen hat er in Fisher (1959) in der ihm eigenen Art nochmals zusammengefaßt. Hinsichtlich der Interpretation der Fisherschen Konzeption siehe vor allem Bartlett (1965) sowie die Beiträge in Fienberg/Hinkley (1980). Herausragend und von Statistikern aller Lager als klassischer Beitrag angesehen ist Savage (1976). Seidenfeld (1979) beschränkt sich größtenteils auf eine Diskussion des Fiduzialarguments. Zu Fishers Ansicht über inverse Wahrscheinlichkeit vgl. Zabell (1989), zu bayesianischen Elementen Barnard (1988). Wie bei anderen führenden Statistikern finden sich auch bei Fisher einige Positionen, die er lediglich einnahm "to avoid agreeing with his opponents", ebenso wie "research for fun". Savage (1976, S. 446, 447).

¹⁹⁴ Vgl. Irwin (1963, S. 166). Nach Savage (1976, S. 449) wurde die Verteilung durch Snedecor nach Fisher F -Verteilung benannt, was ihm Fisher zeitlebens übelnahm.

¹⁹⁵ Dieser Unterschied markiert nach Baird (1983) einen Wendepunkt in der Entwicklung der induktiven Inferenz.

¹⁹⁶ Popper (1994, Anhang *IX) ignoriert z. B. spätere Weiterentwicklungen der Likelihoodinferenz und bringt diese lediglich mit Fisher in Verbindung.

¹⁹⁷ Vgl. Bartlett (1965, S. 396).

exakten Verteilungen von Stichprobenkennzahlen, eine Reihe von Prinzipien statistischer Inferenz und schließlich die Entwicklung allgemeingültiger Richtlinien für den Entwurf von Experimenten.

1922 machte Fisher in einem umfangreichen Beitrag erstmals die klare Trennung zwischen theoretischen Parametern und ihren Statistiken deutlich. Zu dieser Zeit waren Begriffe wie Mittelwert, Standardabweichung etc. sowohl für theoretische Parameter als auch für deren Schätzungen in Gebrauch.¹⁹⁸ Eine Reihe von zentralen Konzepten wurde von Fisher zum ersten Mal behandelt oder zumindest erstmals systematisch eingeordnet: so etwa die Forderung nach Konsistenz, Effizienz und Suffizienz eines Schätzers. Weiterhin zeigte er, daß die bis dahin hauptsächlich verwendeten Statistiken, die von K. Pearson propagierten Momentenschätzer, in der Regel weder effizient noch konsistent und erwartungstreu sind.¹⁹⁹ Statt dessen schlug er die (für einen speziellen Fall bereits 1912 von ihm eingeführte) Maximum-Likelihood-Methode als allgemeines Schätzprinzip vor, da diese Methode automatisch Schätzer liefere, die die genannten Kriterien erfüllen.²⁰⁰

Für Fisher bestand die Aufgabe der Statistik in einer *Reduktion von Daten ohne Informationsverlust*. Er ging von drei fundamentalen Säulen der theoretischen Statistik aus:²⁰¹

- (1) der Modellspezifikation, d. h. der Auswahl der vermuteten (parametrischen) Verteilungsgesetze für ein bestimmtes Problem,²⁰²
- (2) der Schätzung dieser unbekannten Parameter mit einem angemessenen Verfahren²⁰³ und
- (3) der Kenntnis der (exakten oder approximativen) Verteilung dieser Schätzer.

Diese Erkenntnisziele bilden auch heute noch die prinzipiellen Erkenntnisziele der Statistik. Die von Fisher erstmals systematisierten *Kriterien* für eine Schätzung seien kurz erläutert: Zuerst solle ein Schätzer konsistent sein: "That when applied to the whole population the derived statistic should be equal to the parameter."²⁰⁴

¹⁹⁸ Vgl. Fisher (1922, S. 13). Er kritisierte hier (S. 31) explizit Pearson/Filon (1898).

¹⁹⁹ Menges (1972, S. 308): "Nach dem harten Verdikt R. A. Fishers sind die Momentenschätzungen fast ganz aus der Statistik verschwunden. Aber eine solche Vernachlässigung ist eigentlich unberechtigt, denn die Momentenmethode hat einen sehr großen Vorteil. Man braucht nichts über die Verteilungsfunktion der Stichprobe zu wissen. Gerade in den Sozialwissenschaften, wo der Mangel an A-priori-Wissen sehr verbreitet ist, hätte man allen Grund, die Momentenmethode stärker zu beachten."

²⁰⁰ Geisser (1992, S. 9) bezeichnet Fisher (1922) daher zu Recht als "landmark event in theoretical statistics".

²⁰¹ Vgl. Fisher (1922 [1992, S. 15ff]).

²⁰² Diese "Datenreduktion", eine Trennung von relevanter und irrelevanter Information, impliziere auch, daß man nur die Parameter, an denen man interessiert sei, modellieren solle. Das Ziel der statistischen Erkenntnis war für Fisher also ein Ausschnitt aus der Realität, der mit den richtigen statistischen Methoden erfaßt werden könne. Somit ging er davon aus, daß postulierte Modelle zwar nur partiell, aber immerhin "wahr" seien (vgl. dazu unten).

²⁰³ "Estimates are of central importance to Fisher, but I doubt that he attempted any precise definition of the concept." Savage (1976, S. 453).

Diese Forderung müsse nicht notwendigerweise nur ein einziger Schätzer erfüllen. Daher forderte Fisher als weiteres Kriterium, von allen konsistenten Schätzern sei derjenige vorzuziehen, der die geringste Varianz aufweise.²⁰⁵ Dieser werde als "effizient" bezeichnet. Diese Aussage bezieht sich auf unendlich viele, endliche, gleich große Daten (Stichproben), die demselben Verteilungsgesetz unterliegen. Dieses Kriterium sei aber in gewisser Hinsicht unzureichend, "for different methods of calculation may tend to agreement for large samples, and yet differ for all finite samples."²⁰⁶ Daher sei darüber hinaus von einem Schätzer das Kriterium der "Suffizienz"²⁰⁷ zu fordern: Dieser solle die gesamte Information, die in den Daten über den zu schätzenden Parameter enthalten ist, verarbeiten. Des weiteren betonte Fisher, daß die von ihm genannten bzw. geforderten Eigenschaften zwar wünschenswert seien, falls ein Schätzer vorliege, sie aber keine Auskunft darüber geben könnten, welche allgemeine Schätzmethode zu Schätzern führe, die diese Eigenschaften aufweisen. Die bis dahin populäre Methode der Momentenschätzer weise in vielen Situationen eine oder mehrere Eigenschaften nicht auf. Ganz anders jedoch die Maximum-Likelihood-Methode.²⁰⁸ Deren Schätzer basieren auf Größen, die von Fisher "Likelihoods" genannt wurden bzw. im stetigen Fall als Likelihoodfunktion bezeichnet wurde.²⁰⁹ Wir wollen im folgenden nicht auf technische Einzelheiten eingehen, sondern die diesem Ansatz zugrundeliegende Inferenzkonzeption näher betrachten. Fishers

²⁰⁴ Fisher (1922 [1992, S. 18]). Heute findet sich im allgemeinen eine etwas andere Definition: Die Eigenschaft der Konsistenz (eine asymptotische Eigenschaft) besagt in heutiger Diktion, daß der Schätzer für den wahren Parameter θ mit zunehmendem Stichprobenumfang stochastisch gegen den wahren Parameter konvergiert bzw. daß die Wahrscheinlichkeit dafür, daß der Abstand zwischen dem Schätzer und θ für $n \rightarrow \infty$ kleiner als ein beliebiges ε wird, gleich Eins ist.

²⁰⁵ Dieses Konzept beziehe sich auf "large samples, when the distributions of the statistics tend to normality". Dann sei die Statistik zu wählen, die den "least probable error" aufweise (Fisher (1922 [1992, S. 18])). Auf die von Fisher behandelten und zum Teil von Geisser (1992) kommentierten technischen Details gehen wir hier nicht näher ein, so etwa auf das Problem, daß die Größe "Varianz" in einigen Situationen ein ungeeignetes Maß für die Exaktheit eines Schätzers darstellt. Fisher hat diese Frage später mit dem Konzept der "Information" präzisiert.

²⁰⁶ Fisher (1922 [1992, S. 18]).

²⁰⁷ In der deutschsprachigen Literatur zur theoretischen Statistik wird eine "suffiziente" Statistik oftmals auch als "erschöpfende" Statistik bezeichnet.

²⁰⁸ Fisher hat die Allgemeingültigkeit dieser Behauptung nur postuliert, nicht bewiesen.

²⁰⁹ Siehe dazu die Beispiele in Anhang A1. Fisher begründete später (Fisher (1956, S. 15)) die Wahl des Begriffes wie folgt: "Was jetzt klar wurde, ist, daß der mathematische Begriff der Wahrscheinlichkeit in den meisten Fällen ungeeignet ist, unser Vertrauen oder Misstrauen [sic!] zu solchen Schlußfolgerungen auszudrücken und dass die mathematische Größe, die gegeben erscheint, die verschiedenen möglichen Populationen der Güte ihrer Eignung entsprechend, anzuordnen, tatsächlich nicht den Gesetzen der Wahrscheinlichkeit gehorcht. Um diese Größe von "Wahrscheinlichkeit" zu unterscheiden, habe ich den Ausdruck "*Likelihood*" für sie gebraucht, da sowohl "Likelihood" als auch "Probability" (Wahrscheinlichkeit) im gewöhnlichen (englischen) Sprachgebrauch beide Arten der Beziehungen ausdrücken." Hervorhebung im Original. Eine deutsche Bezeichnung hat sich nicht eingebürgert. Anderson hatte ursprünglich geplant, den Begriff in seinem Beitrag für das *Handwörterbuch der Sozialwissenschaften* mit "Wahrheitsähnlichkeit" zu übersetzen, sich dann aber für "Mutmaßlichkeit" entschieden. Vgl. Anderson (1959 [1963, S. 949, nach S. 950]).

Auffassung war zunächst von einer insbesondere gegen Pearson (1920) gerichteten expliziten Ablehnung inverser Wahrscheinlichkeit gekennzeichnet. Diese Sichtweise sei seiner Ansicht nach vor allem durch die Vermengung von theoretischen Parametern und Schätzungen zustande gekommen:

“It is this last confusion, in the writer’s opinion, more than any other which has led to the survival of the present day of the fundamental paradox of inverse probability, which like an impenetrable jungle arrests progress towards precision of statistical concepts.”²¹⁰

Gleichzeitig entwickelte er jedoch ein gewisses Verständnis:

“The criticisms of Boole, Venn, and Chrystal have done something towards banishing the method, at least from the elementary text-books of Algebra; but though we may agree wholly with Chrystal that inverse probability is a mistake (perhaps the only mistake to which the mathematical world has so deeply committed itself), there yet remains the feeling that such a mistake would not have captivated the minds of Laplace and Poisson if there had been nothing in it but error.”²¹¹

Fishers Wahrscheinlichkeitskonzeption war objektiv, doch lehnte er eine Definition von Wahrscheinlichkeit als Grenzwert der relativen Häufigkeit bei einer unbegrenzten Anzahl wiederholter Versuche – die von Misessche Definition, der sich die meisten Objektivisten anschlossen – vehement ab:²¹² “For Fisher, a probability is the fraction of a set, having no distinguishable subsets, that satisfies a given condition [...]”²¹³

Fisher postulierte, daß sich die statistische Inferenz auf theoretische und somit feste Parameter hypothetisch infiniten Populationen zu beziehen habe und markierte damit die Richtung für Forschungen auf dem Gebiet der theoretischen Statistik für die nächsten fünfzig Jahre.²¹⁴ Andererseits konnte sich seine Auffassung einer statistischen oder “wissenschaftlichen” Inferenz nicht durchsetzen. Er verwendete den Ausdruck der “induktiven Logik” nicht zuletzt, um sich von dem Ansatz seines intellektuellen Rivalen J. Neyman abzusetzen, der von einem “inductive behavior” sprach.²¹⁵ Für den Fall, daß eine unstrittige A-priori-Verteilung vorlag, sei es

²¹⁰ Fisher (1922 [1992, S. 13]), ähnlich noch in Fisher (1959, S. 34). Bei Fisher (1956, S. 9) findet sich eine (mehr oder weniger) eindeutige Ablehnung des Bayes-Ansatzes.

²¹¹ Fisher (1922 [1992, S. 13]). Solche Ambiguitäten sind charakteristisch für Fishers Werk. Nach Geisser (1992, S. 4) vertrat Fisher zumindest noch 1912 Ansätze einer Bayes-Logik. Hier (S. 26f) lehnte er die Gültigkeit des Bayes-Theorems explizit ab. Vgl. zu dieser Frage auch Barnard (1988).

²¹² Siehe die Belege in Savage (1976, S. 461). In Fisher (1959, S. 32) betonte er z. B., daß mit jener Definition keine Wahrscheinlichkeiten von Einzelereignissen erfaßt werden könnten.

²¹³ Savage (1976, S. 461) mit entsprechenden Belegen. Savage bemerkt hierzu: “Such a notion is hard to formulate mathematically, and indeed Fisher’s concept of probability remained very unclear, which must have contributed to his isolation from many other statistical theorists.” (S. 462).

²¹⁴ Vgl. Geisser (1992). Bis heute haben sich zum Teil zweideutige Begriffe wie “Mittelwert”, “Standardabweichung” oder “Korrelationskoeffizient” gehalten und bezeichnen in unterschiedlichen Zusammenhängen entweder theoretische Größen oder Schätzer für diese theoretischen Größen.

²¹⁵ Vgl. Savage (1976, S. 462) mit Belegen.

möglich, von Wahrscheinlichkeiten von Ereignissen zu sprechen; diese seien als Fiduzialwahrscheinlichkeiten zu bezeichnen.²¹⁶ Intervalle, die die Unsicherheit einer Schätzung ausdrücken, seien stets als Fiduzialintervalle zu konstruieren. Eng mit dem Problem der Intervalle zur Angabe der Genauigkeit einer Schätzung ist das Problem des "Signifikanztests" verbunden. Was wir heute unter der Logik des Signifikanztests verstehen, gewann in den ersten zwei Jahrzehnten des 20. Jahrhunderts zunehmend an Bedeutung.²¹⁷ Sie geht zum großen Teil auf Fisher zurück und hat sich neben der von Neyman und Pearson entwickelten Hypothesentestkonzeption (s. u.) erhalten. Für Fisher ist das Signifikanzniveau eines Tests ein *Evidenzmaß*, das weder a priori und unabänderlich festgelegt werden sollte noch eine Handlungsrichtlinie definiert:

"A man who 'rejects' a hypothesis provisionally, as a matter of habitual practice, when the significance is at the 1 % level or higher, will certainly be mistaken in not more than 1 % of such decisions. For when the hypothesis is correct he will be mistaken in just 1 % of these cases, and when it is incorrect he will never be mistaken in rejection. This inequality statement can therefore be made. However the calculation is absurdly academic, for in fact no scientific worker has a fixed level of significance at which from year to year, and in all circumstances, he rejects hypotheses; he rather gives his mind to each particular case in the light of his evidence and his ideas."²¹⁸

Diese Kritik war gegen eine Konzeption gerichtet, die seit den dreißiger Jahren von J. Neyman und E. S. Pearson propagiert wurde und sich schnell zur dominierenden Sicht entwickelte. Bevor wir uns der Neyman-Pearson-Theorie zuwenden, wollen wir noch kurz einen Blick auf die parallel verlaufende Entwicklung der (philosophischen) Wahrscheinlichkeitstheorie werfen. Fisher berief sich in seiner Wahrscheinlichkeitskonzeption auf häufigkeitstheoretische Definitionen von G. Boole, J. Venn und G. Chrystal,²¹⁹ vertrat darüber hinaus aber auch einen mengentheoretischen Wahrscheinlichkeitsbegriff. Eine Weiterentwicklung der häufigkeitstheoretischen Positionen fand zwischen etwa 1915 und 1935 vor allem durch Richard von Mises und Hans Reichenbach statt. Von Mises konstatierte rückblickend zur Entwicklung der Wahrscheinlichkeitstheorie:

²¹⁶ Ebda., S. 466: "Nobody knows just what they mean [...]. In a word, Fisher hopes by means of some process – the fiducial argument – to arrive at the equivalent of posterior distributions in a Bayesian argument without the introduction of prior distributions [...]." Wir wollen uns dieser Kritik anschließen. Zwar bemerkt Menges (1972, S. 275): "Das Fiduzialkonzept betrachtet die Ergebnisse einer Beobachtung somit als unumstößliche Tatsache, als die Basis, auf welcher die Inferenz aufzubauen ist. *Es kann damit prinzipiell dem historischen Charakter sozialer Phänomene gerecht werden*" (Hervorhebungen im Original), doch gilt dies unseres Erachtens auch für eine bayesianische Logik.

²¹⁷ So etwa der von Pearson 1900 eingeführte χ^2 -Anpassungstest, "Students" 1908 entwickelter und von Fisher formalisierter *t*-Test oder der von Fisher 1924 eingeführte *F*-Test in der Varianzanalyse.

²¹⁸ Fisher (1959, S. 41f). Die Tatsache, daß Fisher in seinem berühmten Lehrbuch *Statistical Methods for Research Workers* keine *P*-Werte, sondern Signifikanzwerte tabellierte, war darin begründet, daß für erstere K. Pearson das Copyright besaß. Vgl. Watson (1983, S. 714).

²¹⁹ So etwa noch in Fisher (1959, S. 17-35).

“Begonnen hat diese Entwicklung mit dem Mißverständnis einer der klassischen Formeln der Wahrscheinlichkeitsrechnung (der Bayesschen Regel), und sie führte durch eine Periode, in der die abwegige Idee verfolgt wurde, aus kleinen Beobachtungsfolgen statistische Schlüsse zu ziehen: sogenannte ‘Small Sample Theory’. Es scheint, daß jetzt die Irrtümer im wesentlichen überwunden sind und einer sachgemäßen Weiterbildung der Formeln, die für eine extensive Anwendung statistischer Methoden nützlich sind, Platz gemacht haben.”²²⁰

Die Theorie von Mises basierte auf dem Begriff des Kollektivs. Ein Kollektiv ist eine Folge von Einzelbeobachtungen gleichartigen Charakters, bei denen die relative Häufigkeit des Auftretens eines Merkmals bei einer unbegrenzten Anzahl von Wiederholungen einem Grenzwert zustrebt.²²¹ Dieser Grenzwert wurde von ihm als Wahrscheinlichkeit definiert. Gegenüber älteren Häufigkeitstheorien zeichnete sich diejenige von Mises durch eine Ergänzung aus. Demnach sollte der Begriff der Wahrscheinlichkeit auch auf einem Zufälligkeitssaxiom beruhen. Eine wichtige Folgerung aus dieser Konzeption war: “Demnach ist es für von Mises und seine Anhänger ganz unmöglich, etwa über einen konkreten Würfel und die Wahrscheinlichkeit des Würfels einer seiner Seiten zu urteilen [...]”²²²

Eine Position, die sich weder mit der Laplaceschen Konzeption einer inversen Wahrscheinlichkeit noch mit der von von Mises und Reichenbach weiterentwickelten häufigkeitstheoretischen Definition in Einklang bringen ließ, sondern stattdessen einen logischen Wahrscheinlichkeitsbegriff postulierte, wurde 1921 von J. M. Keynes präsentiert. Allerdings hat dieser Ansatz weder in der Statistik noch in der Ökonometrie eine Inferenzkonzeption begründet,²²³ so daß wir hier nicht weiter darauf eingehen, sondern uns nun der mittlerweile als “klassisch” bezeichneten Konzeption von J. Neyman und E. S. Pearson zuwenden wollen.

e. Grundlegung der “klassischen” Sicht: J. Neyman und E. S. Pearson

Als Meilenstein der theoretischen Statistik werden einhellig die Arbeiten von J. Neyman²²⁴ und E. S. Pearson²²⁵ angesehen. Während Fisher in bezug auf das

²²⁰ Von Mises (1951, S. Vf).

²²¹ Vgl. Scherff (1980, S. 571).

²²² Ebda., S. 572. In dieser Strenge folgten von Mises nicht alle Häufigkeitstheoretiker. Für Weiterentwicklungen siehe ebda. sowie Howson (1995, S. 14).

²²³ Vgl. zu diesem Ansatz ausführlich Conniffe (1992).

²²⁴ Jerzy Neyman [1894-1981] war polnischer Herkunft. Er studierte Mathematik an der Universität von Kharkov (Rußland) und verbrachte als Pole in Rußland nach dem Rigaer Frieden 1921 mehrere Monate in Haft. Die erneute Drohung einer Verhaftung zwang ihn anschließend, sein gerade begonnenes Physikstudium abzubrechen. Nach erfolglosen Bewerbungen als Mathematiker an mehreren polnischen Universitäten nahm er Stellungen am staatlichen Institut für Meteorologie sowie als Statistiker an einem Institut für Agrarstudien im deutschen Teil Polens an. Anfang der dreißiger Jahre diente er dem “Institute for Social Problems” als Berater für “survey sampling” und ging 1934 nach kurzem Aufenthalt in Berlin an das *University College* nach London, wo er K. Pearson und W. Gosset traf. Dort begann auch bereits einige Jahre früher seine Zusammenarbeit mit E. S. Pearson. R. A. Fisher, der in diesen Jahren ebenfalls am *University College* lehrte, entwickelte gerade erste Ansätze seiner Fiduzialtheorie. Während Neyman anfangs seine Theorie der Konfidenzintervalle im Einklang mit Fishers Fiduzialintervallen sah, entpuppte sich bald ihr grundlegender Gegen-

Testen von Hypothesen nur die Alternativen "Ablehnung" und "keine Aussage möglich" gelten lassen wollte, entwickelten Neyman und Pearson eine geschlossene Testtheorie, die Ablehnungs- und Annahmebereiche unterschied und Begriffe wie die "Macht" eines Tests, "Fehler erster und zweiter Art" und "gleichmäßig bester Test" einführt. Bis zum Ende des 19. Jahrhunderts basierte das Testen von Hypothesen auf Verteilungen von Teststatistiken, die (1) bestenfalls für große Stichproben angemessen waren und (2) aus intuitiven Gründen herangezogen wurden. Durch die Einführung der t -Verteilung durch W. S. Gosset (1908) und die Beiträge R. A. Fishers, der die exakten Verteilungen von t , χ^2 , F und einiger Korrelationskoeffizienten in Normalverteilungen ableitete, konnte zumindest Problem (1) überwunden werden. Mit der Lösung dieses Problems stellte sich dann die Frage nach einer formal befriedigenden Testtheorie. E. S. Pearson berichtete in einer Rückschau, die Idee zu dieser Theorie durch einen Hinweis von Gosset bekommen zu haben:

"I had been trying to discover some principle beyond that of practical expediency which would justify the use of "Student's" ratio $z = (x-m)/s$ in testing the hypothesis that the mean of the sample population was at m . Gosset's reply (to the letter in which Pearson [...] had raised the question) had a tremendous influence on the direction of my subsequent work, for the first paragraph contains the germ of that idea which has formed the basis of all the later joint researches of Neyman and myself. *It is the simple suggestion that the only valid reason for rejecting a statistical hypothesis is that some alternative hypothesis explains the observed events with a greater degree of probability.*"²²⁶

Gosset argumentierte in diesem Brief, daß ein noch so kleiner Wahrscheinlichkeitswert wie etwa 0.0001 für eine Zufallsstichprobe per se nicht zu einer Ablehnung einer Hypothese führen könne. Lediglich der Vergleich mit einer Alternativhypothese, "which will explain the occurrence of the sample with a more reasonable probability, say .05 (such as that it belongs to a different population or that the sample wasn't random or whatever will do the trick) you will be very much more inclined to consider that the original hypothesis is not true."²²⁷

satz und bildete den Ausgangspunkt für erbitterte Auseinandersetzungen zwischen den beiden Wissenschaftlern. 1938 ging Neyman nach einem kurzen Gastaufenthalt 1937 in die USA, nahm einen Ruf als Professor für Statistik an der Universität von Berkely an und wurde Direktor des dortigen "Statistical Laboratory", wo er bis zu seinem Tode 1981 blieb. Die von ihm alleine und zusammen mit E. S. Pearson entwickelten Ansätze versuchte er stets über die Theorie hinaus in praktische Anwendungen einzubinden, etwa in den Bereichen der Medizin, Biometrie, Populationsdynamik, im Versicherungswesen, in der Astronomie, Meteorologie, in Zusammenhang mit Hypothesen über Luftverschmutzung, Sicherheit von Kernkraftwerken usw.

²²⁵ Egon S. Pearson [1895-1980] war der Sohn von K. Pearson. An dessen Hauptwirkungsstätte, dem *University College* in London, verbrachte auch E. S. Pearson den Großteil seiner wissenschaftlichen Karriere. Das *Department of Applied Statistics* wurde nach K. Pearsons Emeritierung aufgeteilt. R. A. Fisher wurde dessen unmittelbarer Nachfolger, und E. S. Pearson wurde *Reader* an einem neu eingerichteten *Department of Statistics*. Von 1936 bis 1965 war er Herausgeber der *Biometrika*. Die Entwicklung der Neyman-Pearson-Theorie (die "Neyman-Pearson story") wird instruktiv in Pearson (1966) geschildert.

²²⁶ Pearson in einer Arbeit von 1939, zitiert nach dem Kommentar von Lehmann (1992, S. 68) zu Neyman/Pearson (1933). Hervorhebung von uns.

Diese Idee wurde dann gemeinsam von Neyman und Pearson 1928 in einer umfangreichen zweiteiligen Arbeit in der *Biometrika* zum Konzept des Likelihood-Quotiententests ausgebaut. Während Pearson darin bereits die einheitliche Methode sah, nach der sie gesucht hatten, zeigte sich Neyman noch nicht zufrieden:

“It seemed to him that the likelihood ratio principle itself was somewhat ad hoc and was lacking a fully logical basis. His search for a firmer foundation, which constitutes the third of the three steps, eventually led him to a new formulation: The most desirable test would be obtained by maximizing the power of the test, subject to the condition that under the hypothesis, the rejection probability has a preassigned value, the level of a test.”²²⁸

Das Ergebnis war der Beitrag von 1933, in dem sich auch das berühmte Neyman-Pearson-Lemma findet, das besagt, daß in der Klasse aller Tests mit Sicherheitswahrscheinlichkeit α die Gütefunktion des Likelihood-Quotiententests die Gütefunktion jedes anderen Tests dominiert (d. h. jeder andere Test hat eine größere Fehlerwahrscheinlichkeit zweiter Art). Neyman und Pearson zeigten anhand einer Reihe von Beispielen die Anwendung dieses Prinzips und legten damit den Grundstein für eine weithin anerkannte allgemeine Testtheorie, die heute zusammen mit den ebenfalls von Neymann formulierten “Konfidenz”intervallen als “klassisch” bezeichnet wird. Die Vorgehensweise nach der Neyman-Pearson-Logik läßt sich nach Lehmann durch vier Schritte beschreiben:²²⁹

- (1) Spezifikation eines Modells aus einer parametrischen Familie von Verteilungen, das die Daten hervorgebracht hat,
- (2) Spezifikation einer Hypothese bezüglich eines interessierenden Parameters: $H_0: \theta = \theta_0$ und einer einfachen bzw. einer Klasse von Alternativen H_1 , z. B. $\theta \leq \theta_0$,
- (3) Spezifikation eines Signifikanzniveaus α , das die maximal zulässige Wahrscheinlichkeit eines Fehlers erster Art angibt,
- (4) Auswahl der optimalen Prozedur zum Test von H_0 gegen H_1 durch Minimierung des β -Fehlers.²³⁰
Lehmann führt schließlich noch einen fünften Punkt an, der aber eher eine Voraussetzung als eine Vorgehensweise darstellt:
- (5) Alle (vier) Schritte müssen vollzogen sein, bevor man die Daten betrachtet (“before any observations have been seen”).

²²⁷ Ebda.

²²⁸ Lehmann (1992, S. 68). Dieser wichtigste Aspekt der Neyman-Pearson-Theorie wird oftmals nicht beachtet. So betont Borovcnik (1992, S. 92) zu Recht, daß “[...] eine Häufigkeitsdeutung des α -Fehlers beim Testen überstrapaziert [wird], während der eigentliche Witz des Verfahrens in der Minimierung des β -Fehlers liegt.”

²²⁹ Nach Lehmann (1992, S. 69f).

²³⁰ Auf die entsprechenden Techniken wollen wir hier nicht eingehen und statt dessen auf die Lehrbuchliteratur verweisen. Ein Standardwerk ist Lehmann (1964).

Das von Neyman und Pearson postulierte Vorgehen war im eigentlichen Sinne nichts anderes als eine Verhaltensrichtlinie. Die beiden Autoren formulierten die hinter ihrer Theorie stehende Überzeugung wie folgt:

“Without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behaviour with regard to them, in following which we insure that, in the long run of experience, we shall not be too often wrong.”²³¹

Inferenzaussagen sind demnach hypothetisch-deduktiv und nur möglich, *bevor* sich Ereignisse realisieren. Sie beziehen sich damit nicht auf spezifische Hypothesen, sondern auf die “lange Sicht”, auf zukünftige Handlungen. Konsequenterweise wurde dieser Ansatz später von A. Wald zu einer reinen Entscheidungstheorie erweitert, und auch Neyman hat in seinen späteren Arbeiten immer wieder diesen verhaltenstheoretischen Aspekt betont.²³²

Vehemente Kritik kam dagegen von keinem geringeren als R. A. Fisher, der die Neyman-Pearson-Theorie eventuell für Situationen, in denen permanent Entscheidungen getroffen werden mußten, anerkennen wollte, auf keinen Fall aber für inferenzstatistische Beurteilungen in einem *wissenschaftlichen* Sinne. Ein weiteres Argument betraf die Forderung eines “repeated sampling from the same population”. In Anlehnung an J. Venn wies Fisher darauf hin, daß eine vorliegende Stichprobe stets aus einer Vielzahl verschiedener denkbarer Populationen hervorgegangen sein könnte: “so [...] the phrase ‘repeated sampling from the same population’ does not enable us to determine which population is to be used to define the probability level, for no one of them has objective reality, all being products of the statistician’s imagination.”²³³

Auch von anderer Seite wurden dieser Vorgehensweise Vorbehalte entgegengebracht: Zum einen würden *in der Praxis* Modelle meist im Lichte der Daten ausgewählt, weiterhin würden oftmals nicht nur eine, sondern mehrere Hypothesen anhand derselben Daten untersucht, und schließlich sei in vielen Situationen die Reduktion der Inferenz auf eine Ja/Nein-Entscheidung nicht angemessen.

Weiterhin hat sich auch gezeigt, daß optimale (im Sinne gleichmäßig mächtigster) Tests nur für eingeschränkte Situationen existieren oder aber (im Sinne der Maximierung der minimalen Macht) so komplex sind, daß ihre Anwendung erhebliche Probleme bereitet. Es ist aber zu betonen, daß diese Vorbehalte die Ausnahme darstellen und die überwiegende Mehrzahl, vor allem im Bereich der angewandten Statistik, den Neyman-Pearson-Ansatz vorbehaltlos akzeptiert hat, er sich geradezu zu einem Paradigma entwickelt hat, auch wenn sich die Statistiker bis heute darüber streiten, wo denn nun genau die Unterschiede zwischen diesem Ansatz und der Fisherschen Testkonzeption liegen.²³⁴

²³¹ Neyman/Pearson (1933 [1992, S. 74]). Kyburg (1985, S. 119) bringt die Intention auf den Punkt: “That says nothing about the case before us, but it may make us feel better.”

²³² Siehe z. B. Neyman (1971).

²³³ Fisher (1955, S. 71).

²³⁴ Vgl. etwa Lehmann (1993). Für “Bayesianer” scheint dieser Unterschied von noch geringerer Bedeutung zu sein. So spricht etwa Barlow (1992, S. 128) in seinem Kommentar zu de Finetti (1937) von einem ‘Fisher-Neyman-Pearson’-Ansatz: “The ‘Fisher-Neyman-Pearson’ ap-

Vergleicht man diesen Ansatz mit demjenigen Fishers, so ist insbesondere der angeführte 5. Punkt entscheidend. Demnach ist das Vorgehen nach Neyman und Pearson streng deduktiv, während Fishers Ansatz (zumindest auch) induktiv ist, die Beurteilung erst nach der Evidenz der Daten geschehen sollte; und dies vor allem nicht in Hinblick auf eine Alternativhypothese. Sicher war es nicht Neymans und Pearsons Intention, ein universales, konstantes Signifikanzniveau zu propagieren, nur: Auch wenn sie unterschiedliche Niveaus für unterschiedliche Situationen zulassen, so müssen diese vor dem Experiment bzw. vor der Kenntnis der Evidenz der Daten festgelegt worden sein. Der zweite fundamentale Unterschied besteht in der *Richtung* der Inferenz. Fishers und die in dieser Hinsicht logisch äquivalente Signifikanztestkonzeption K. Pearsons ist auf einen Zustand gerichtet, der besteht oder strenggenommen auch schon vergangen sein kann. Neymans und Pearsons Inferenz ist dagegen in die Zukunft gerichtet: Wenn wir aufgrund des Tests zukünftig auf diese oder jene Art handeln werden, wie oft werden wir dann einen Fehler machen? In der Praxis ist, so bleibt festzuhalten, gegenwärtig eine Vermengung beider Konzepte festzustellen.²³⁵

Eine Alternative zu diesen Signifikanz- und Hypothesentestkonzeptionen ergab sich aus einer Weiterentwicklung der Likelihood-Ideen von R. A. Fisher, die vor allem von G. A. Barnard geleistet und dann von A. Birnbaum zu einem Likelihood-*Prinzip* ausgebaut wurde.

Bevor wir uns dieser Theorie zuwenden, wollen wir diesen Abschnitt mit einer Bemerkung zu der von A. Wald entwickelten sog. "Entscheidungstheorie" abschließen, in der viele die logische Weiterentwicklung der Neyman-Pearson-Theorie zu einer Neyman-Pearson-Wald-Theorie (NPW-Theorie) sehen.²³⁶ In Anlehnung an Stegmüller wollen wir diese Sicht ablehnen, die Inferenzstatistik als *Schätztheorie* eindeutig von der Entscheidungstheorie als *Schätzhandlung* trennen und somit letztere als für eine theoretische Erörterung der inferenzstatistischen Grundlagen nicht relevante Problematik außer acht lassen.²³⁷ Es muß allerdings

proach became dominant in the United States in the field of statistics." Im Vergleich zu de Finettis Position ist der Unterschied zwischen diesen beiden Schulen in der Tat unerheblich (s. u.). Eine Reihe von Kontroversen zwischen Fisher und Neyman/Pearson ist nach Savage (1976, S. 448) auf mangelnde Objektivität zurückzuführen: "In science, it is hostility rather than familiarity that breeds contempt, and all of Fisher's castigation of the Neyman-Pearson school [hier folgen zahlreiche Belege, T. R.] shows that he never had sufficient respect for the work of that school to read it attentively [...]. And members of that school in referring to Fisher were likely to read their own ideas impatiently into his lines."

²³⁵ Johnstone (1986, S. 6) beschreibt die herrschende Praxis treffend: "In general, tests of significance in practice follow Neyman formally, but Fisher philosophically. Formally, there is mention of 'alternative' hypotheses, errors 'of the second kind', and the 'power' of the test, which are terms due to Neyman (and his colleague Pearson). But philosophically, the result in a test, e.g. the result that the level of significance P equals 0.049, or that P is less than or equal to 5 %, is interpreted as a measure of evidence, which is the interpretation following Fisher, and denied repeatedly by Neyman."

²³⁶ Vgl. etwa Wald (1950). Wald verunglückte 1950 tödlich und konnte die Diskussion um seine Theorie nicht mehr verfolgen. Vgl. zur Bedeutung und Rezeption der Entscheidungstheorie in der Statistik die Hinweise von Weiss (1992), zu historischen Vorläufern Bott (1962).

²³⁷ Stegmüller (1973, S. 194ff). Er betont unseres Erachtens zu Recht: "Theoretische Schätzungen gehören zur Begriffsfamilie der Überzeugungen [im Gegensatz zu Schätzhandlungen

betont werden, daß die hier vertretene Ablehnung einer grundsätzlichen Bedeutung der Entscheidungstheorie vermutlich mindestens ebenso viele Argumente gegen wie für sich hat.²³⁸

f. Inferenz für gegebene Daten und Hypothesenwahrscheinlichkeit

Auch wenn sich die Ansätze von Fisher und Neyman / Pearson in mancher Hinsicht unterscheiden, so ist ihnen doch im Vergleich zu den nun zu besprechenden Alternativen eines gemeinsam: Beide basieren auf einer Ablehnung bayesianischer Sichtweisen, die dann aufgrund ihrer Arbeiten von der großen Mehrheit der Wissenschaftler geteilt wurde.²³⁹

Dennoch entwickelte sich parallel eine moderne bayesianische Inferenzstatistik. Diese Entwicklung vollzog sich in drei Schritten: einer neuen Begründung bayesianischer Wahrscheinlichkeitstheorien²⁴⁰ durch F. P. Ramsey, B. de Finetti, H.

der Entscheidungstheorie, T. R.). Und *Überzeugungen haben überhaupt keine praktischen Konsequenzen*, jedenfalls nicht in dem Sinn, in welchem Handlungen Konsequenzen besitzen." (S. 195, Hervorhebungen im Original). In ähnlichem Sinne argumentierte bereits Anderson (1957, S. 943), der ihre wissenschaftliche Relevanz (in Anlehnung an R. A. Fisher) kritisch beurteilte: "Ob dieselbe Neyman-Pearson-Waldsche Theorie auch auf die Wahl bzw. Prüfung von *wissenschaftlichen* Hypothesen angewandt werden darf, bleibt strittig." (Hervorhebung im Original).

²³⁸ Wie schwierig eine endgültige Klärung dieser Frage ist, belegt etwa die Begründung von Menges (1972, S. VII), weshalb er der Entscheidungstheorie in der Neuauflage seines Lehrbuches mehr Bedeutung als in der ersten Auflage beimißt, die Schätz- und Testtheorie ihr aber (noch) nicht unterordnen will: "Für eine rein entscheidungstheoretische Behandlung der Schätz- und Testtheorie ist die Zeit allerdings noch nicht reif. In der zeitgenössischen Statistik dominieren noch immer die klassischen Konzepte. Das wird wohl auch eine Weile noch so bleiben, da die Lösung der Anwendungsprobleme der statistischen Entscheidungstheorie eher schleppend vor sich geht." Noch Schlittgen (1996, S. 4) stellt aber fest, daß es "[...] um die Entscheidungstheorie recht ruhig geworden ist." Einige Vertreter personalistischer Wahrscheinlichkeitstheorien, wie etwa de Finetti, lehnen ebenfalls eine Unterordnung der Schätz- unter die Entscheidungstheorie ab, andere jedoch nicht.

²³⁹ Bjornstad (1992, S. 462f): "When [Fisher (1922)] appeared, the Bayesian theory of Laplace was the main approach to statistical inference. Fisher's likelihood-based alternative, together with his sharp criticism of the use of prior distributions, especially priors to represent ignorance [...], led to a lesser interest in Bayesian inference."

²⁴⁰ Wir wollen hier kurz auf die in letzter Zeit oftmals als Konkurrenz zur Wahrscheinlichkeitstheorie auftauchende "Fuzzy-Logik" eingehen, um Verwechslungen vorzubeugen. Fuzzy-Logik unterscheidet sich laut ihren Verfechtern von der herkömmlichen Logik vor allem dadurch, daß "Zwischenwerte" zugelassen werden. In bezug auf die (klassische) Mengenlehre bedeutet dies etwa, daß ein Element nicht eindeutig einer Menge zugehört oder nicht, sondern zu einem gewissen Grad in einer Menge enthalten ist. Das mittlerweile klassische Beispiel von Zadeh (1965) definiert den Begriff "Größe" als unscharfe Menge. Für den Ausdruck "Person A ist groß zum Grad 0.8" gibt es nach Dröscher (1994, S. 48) folgende Interpretationsmöglichkeiten: "80 % aller Leute bezeichnen Person A als groß", "Person A ist in 80 % der Fälle [?] groß" oder "Die Größe der Person A wird auf x cm geschätzt, da alle Personen ab x cm als groß bezeichnet werden, ist Person A groß, allerdings mit einer Fehlerwahrscheinlichkeit von 20 % [klassischer Signifikanztest?], mithin schließlich: Person A ist "ziemlich groß". Auf einer Skala von 0 bis 1 weist sie den Wert 0.8 auf. Eine andere Interpretation könnte jedoch lauten: Man normiert die Größen aller Personen auf den Bereich [0,1]. Dann ist Person A 0.8 groß. Groß ist nichts anderes als eine Aggregation. Im Fuzzy-Sinne mit mehr Ausprägungen als herkömmlich, aber dennoch wird Information verschenkt. In dem Bereich, in dem die Funktion nicht 0 oder 1 ist, wird die Information der

Jeffreys und L. J. Savage, die Erweiterung der verschiedenen Likelihood-Ansätze zu einem Likelihood-Prinzip durch G. A. Barnard und vor allem A. Birnbaum und schließlich die Verbindung dieser beiden Komponenten zu einer modernen bayesianischen Inferenz, die allerdings zahlreiche Ausprägungen erfahren hat. Betrachten wir zunächst die Entwicklung der subjektivistischen Wahrscheinlichkeitstheorien. Diese gehen nach Howson (1995, S. 2) von den folgenden drei Grundannahmen aus:²⁴¹

1. Eine Hypothese A ist in Extremfällen mit Sicherheit wahr oder mit Sicherheit falsch. Dazwischen sind Grade der Überzeugung (degrees of belief) von A zugelassen.
2. Diese Grade der Überzeugung können numerisch ausgedrückt werden.
3. Wenn sie rational sind und auf dem geschlossenen Einheitsintervall gemessen werden, erfüllen sie die endlichen Additivitätsaxiome.²⁴²

Die subjektivistischen bayesianischen Konzeptionen Ramseys, de Finettis, Jeffreys und Savages sind nacheinander, aber unabhängig voneinander entstanden.²⁴³ Wir wollen sie in chronologischer Reihenfolge kurz ansprechen.

Die erste "moderne" subjektivistische Wahrscheinlichkeitstheorie stammt von F. P. Ramsey.²⁴⁴ Dessen Abhandlungen, die 1926 und 1928 verfaßt wurden, sind erst posthum 1931 veröffentlicht worden.²⁴⁵ Wir hatten gesehen, daß die epistemolo-

Größe nicht 1 zu 1 übertragen. Epistemologisch ergeben sich somit aus der Fuzzy-Logik unserer Ansicht nach keine neuen Aspekte.

²⁴¹ Die Unterscheidung von Menges (1965) zwischen intuitiven Subjektivisten und diese kritisierende "Bayesianer" ist irreführend, da er de Finetti (1937) beiden Gruppen zuordnet. Darüber hinaus ist seine Behauptung, daß für Subjektivisten die Praktikabilität oberstes Kriterium sei, wohl kaum zutreffend.

²⁴² Die Wahrscheinlichkeitsaxiome können dabei entweder über Rationalitätsprinzipien, die sich durch (faire) Wettquotienten ausdrücken lassen, gewonnen werden oder über eine Nutzentheorie. Zu den Einzelheiten vgl. Howson (1995, S. 2ff). Die sog. "Logiker" wie Keynes (1921) oder Carnap (1950) haben keine Inferenzkonzeption begründet und seien daher hier nur erwähnt.

²⁴³ Ein Grund hierfür ist sicher darin zu sehen, daß diese Entwicklungen nicht in den führenden, von der englischen biometrischen Tradition beherrschten Publikationsorganen stattfanden und daher nur auf Umwegen und mit z. T. mehrjähriger Verspätung zur Kenntnis der entsprechenden Wissenschaftler gelangten. De Finettis grundlegender Beitrag wurde darüber hinaus erst 1963 ins Englische übersetzt.

²⁴⁴ Frank Plumpton Ramsey [1903-1930] verbrachte sein kurzes Leben in Cambridge: zuerst als Schüler am *Trinity College*, dann als *fellow* am *King's College*, schließlich als *lecturer* in Mathematik an der Universität. Sein Interesse galt neben der Mathematik hauptsächlich der Philosophie. Er war einer der ersten, der die Bedeutung von Ludwig Wittgensteins *Tractatus logico-philosophicus* erkannte, und seine eigenen publizierten Arbeiten stehen in dessen Tradition. Diese widmeten sich dem Gebiet der Grundlagen der Mathematik und der formalen Logik, aber es finden sich auch zwei ökonomische Abhandlungen darunter, die im *Economic Journal* publiziert wurden: "A contribution to the theory of taxation" und "A mathematical theory of saving". Letztere Arbeit bezeichnete der damalige Herausgeber der Zeitschrift, J. M. Keynes, in einem Nachruf als "one of the most remarkable contributions to mathematical economics ever made". Keynes im *Economic Journal* 1930, zitiert nach der Einleitung von R. B. Braithwaite in Braithwaite (1931, S. xi).

gische Definition der Wahrscheinlichkeit von Bernoulli bis Laplace explizit, bei Gauss, Galton oder Edgeworth implizit, subjektiver Natur war: "Wahrscheinlichkeit" wurde bereits von C. Huygens im Sinne von Wettquotienten gedeutet, Zufall wurde als Unwissenheit definiert, das Prinzip des unzureichenden Grundes implizierte eine A-priori-Gleichverteilung, die über das Bayes-Theorem mit der Evidenz der Daten zu einer A-posteriori-Wahrscheinlichkeit für einen Parameterwert verbunden wurde.²⁴⁶

Ramsey argumentierte in ähnlicher Richtung, verband dies jedoch mit einer Kritik an der logischen und häufigkeitstheoretischen Interpretation. Er rezipierte zunächst einen Ansatz von Keynes. Für Keynes (1921) bedeutete Wahrscheinlichkeit eine logische Beziehung zwischen zwei verschiedenen Sätzen von Propositionen, wobei diese Beziehung über einen "Grad des Glaubens" miteinander verbunden sind:

"Let our premisses consist of any set of propositions h and our conclusion consist of any set of propositions a , then if a knowledge of h justifies a rational degree of belief in a of degree A , we may say that there is a probability-relation of degree A between a and h ."²⁴⁷

Allerdings setzte Keynes nicht voraus, daß alle Grade des Glaubens numerisch meßbar und vergleichbar seien und umging somit große Schwierigkeiten. Ramsey forderte statt dessen, daß Wahrscheinlichkeiten als Wettquotienten ausgedrückt werden und diese rational (konsistent, kohärent) sein müßten.²⁴⁸ Die Ausführungen Ramseys waren jedoch rein philosophischer Natur. Eine *Inferenz*konzeption stellten sie nicht dar. Diese wurde 1937 in einer berühmten Arbeit von Bruno de Finetti geliefert.²⁴⁹ Auch für de Finetti stand fest, daß die Grundlage jeder Wahrscheinlichkeit subjektiver Natur ist.²⁵⁰ Zentral war hier das Bayes-Theorem: Subjektive Einschätzungen / Wahrscheinlichkeiten seien aufgrund der Kenntnis von Daten mittels des Bayes-Theorems stets zu revidieren. Das bedeutete, daß sich subjektivistische Wahrscheinlichkeiten mit zunehmender Evidenz relativen Häufigkeiten annähern.²⁵¹ De Finetti warf der klassischen Statistik keine falschen Ergebnisse, sondern falsche Grundlagen vor:

²⁴⁵ Ramsey (1931a), Ramsey (1931b).

²⁴⁶ Siehe oben, Abschnitt A 3 b.

²⁴⁷ Keynes (1921, S. 4), zitiert nach Kyburg/Smokler (1964, S. 9).

²⁴⁸ Dies bedeutet unter anderem, daß bei einem Ereignis mit subjektiver Wahrscheinlichkeit p dem komplementären Ereignis die Wahrscheinlichkeit $1-p$ zugesprochen wird.

²⁴⁹ Bruno de Finetti [1906-1985] studierte in Mailand und promovierte dort über ein Thema aus dem Bereich der affinen Geometrie. Bis 1931 arbeitete er im mathematischen Büro des italienischen statistischen Zentralamtes (ISTAT), von 1931 bis 1946 bei einer Versicherungsgesellschaft in Triest. Ab 1946 lehrte er an der Universität von Triest, ab 1954 an der ökonomischen Fakultät in Rom, ab 1961 an der dortigen 'faculty of sciences' als Professor für Wahrscheinlichkeitstheorie. W. Stegmüller, Vertreter einer objektiven Wahrscheinlichkeitsinterpretation, ist der Ansicht, daß in der Begründung de Finettis "[...] nicht nur die Wurzel dafür liegen dürfte, daß der Subjektivismus eine so starke Beachtung gefunden hat, sondern [...] außerdem dazu führte, daß viele Wahrscheinlichkeitstheoretiker und Statistiker die subjektive Interpretation akzeptierten und in allen Varianten des Objektivismus nicht mehr zu erblicken bereit sind als verschiedene Spielarten eines zum Scheitern verurteilten metaphysischen Abenteuers." Stegmüller (1973, S. 364).

²⁵⁰ Vgl. de Finetti (1937 [1992]).

“Der überwiegende Teil der modernen Statistik ist in der Praxis vollkommen normal; ihre Grundlagen sind jedoch falsch. Die Intuition hat den Statistiker jedoch vor Fehlern bewahrt. Meine These ist, daß die Bayes’sche Methode das rechtfertigt, was er immer getan hat [...]”²⁵²

In ähnliche Richtung ging 1939 Harold Jeffreys. Er verband eine Wahrscheinlichkeitstheorie mit einer Theorie der Induktion.²⁵³ Jeffreys betonte (wie auch de Finetti), daß ein fundamentales Problem der Wissenschaft darin bestehe, aus Erfahrung zu lernen:

“Knowledge obtained in this way is partly merely description of what we have already observed, but part consists of making inferences from past experience to predict future experience. This part may be called generalization or induction. It is the most important part; events that are merely described and have no apparent relation to others may as well be forgotten, and in fact usually are.”²⁵⁴

Die genaue Vorgehensweise hat Jeffreys in einem System von fünf Regeln zusammengefaßt, die den Prozeß der Induktion konstituieren sollen:²⁵⁵

1. Alle verwendeten Hypothesen müssen explizit genannt werden, und alle Schlußfolgerungen müssen aus diesen Hypothesen folgen.
2. Eine Theorie der Induktion muß folgerichtig sein (self-consistent): Es darf nicht möglich sein, aus Postulaten und beobachteten Daten sich widersprechende Schlüsse zu ziehen.
3. Jede aufgestellte Regel muß praktisch anwendbar sein. Jede Definition muß so gestaltet sein, daß die Schätzung einer Größe durch die Definition erfaßt werden kann.
4. Eine Theorie der Induktion muß den Fall vorsehen, daß sich Inferenzschlüsse als falsch erweisen können.
5. Eine Theorie der Induktion darf keinen empirischen Satz (proposition) a priori ausschließen. Jeder eindeutig formulierte empirische Satz muß im Sinne von Regel 4 akzeptiert werden können, wenn eine angemessene (moderate) empirische (relevante) Evidenz vorhanden ist.

Regel 1 und 2 sind mathematischer Natur. Regel 3 und 5 unterscheiden empirische und A-priori-Sätze. Regel 4 bedeutet, daß wissenschaftliche Gesetze modifiziert oder möglicherweise ersetzt werden müssen, wenn neue Daten hinzugenommen

²⁵¹ Siehe dazu Beispiel 2 in Anhang A1.

²⁵² De Finetti (1981, S. 657).

²⁵³ Sir [seit 1953] Harold Jeffreys [1891-1989] studierte in Cambridge, wurde dort 1914 zum *fellow* gewählt (mit der Dauer seiner *fellowship* von 75 Jahren hält er den Rekord in Cambridge), verbrachte dort seine gesamte Laufbahn als Professor für Astronomie und hat sich vor allem als Geophysiker (somit durch die Analyse von passiven Beobachtungen) weltweit einen Namen gemacht. Jeffreys ist der einzige “Bayesianer”, mit dem sich Popper (1994, v. a. Anhang *V, *VII, *VIII, *IX) explizit auseinandergesetzt hat.

²⁵⁴ Jeffreys (1939, S. 8).

²⁵⁵ Jeffreys (1961). Das folgende nach Zellner (1971, S. 7-11).

werden. Zusätzlich zu diesen fünf Regeln formulierte Jeffreys drei Richtlinien für wissenschaftliche Forschungen:

6. Die Anzahl der Postulate sollte minimal sein (Einfachheitspostulat).
7. Die Theorie sollte dem Verlauf der menschlichen Gedankengänge folgen.²⁵⁶
8. Ein Einwand (objection) hat kein Gewicht, wenn ein gleichwertiger Einwand einen Teil der allgemein akzeptierten reinen Mathematik bzw. dessen Regeln verletzt.

Diese Regeln haben einige interessante Implikationen. Regel 3 z. B. schließt die klassische Wahrscheinlichkeitsdefinition als Anteilswert in einer unendlichen Folge von Beobachtungen aus:

“They rule out [...] any definition of probability that attempts to define probability in terms of infinite sets of possible observations, for we cannot in practice make an infinite number of observations. The Venn limit, the hypothetical infinite population of Fisher [...] are useless to us by rule 3 [...]. In fact no ‘objective’ definition of probability in terms of actual or possible observations, or possible properties of the world, is admissible. For, if we made anything in our fundamental principles depend on observations or on the structure of the world, we should have to say either (1) that the observations we can make, and the structure of the world, are initially unknown; that we cannot know our fundamental principles, and we have no possible starting point; or (2) that we know a priori something about observations on the structure of the world, and this is illegitimate by rule 5.”²⁵⁷

Wahrscheinlichkeit, so folgt daraus, ist keine Häufigkeit, sondern ein “reasonable degree of belief, which satisfies certain rules of consistency and can in consequence of these rules be formally expressed by numbers.”²⁵⁸ Wenn eine Erklärung für ein beobachtetes Ereignis gegeben werde, könne ein Forscher feststellen, daß diese “probably true” sei. Damit sei impliziert, daß er einen hohen Grad des Vertrauens in eine Hypothese besitze, der wiederum 1. quantifizierbar sei und 2. auf Erfahrung und Information beruhe.²⁵⁹ Regel 7 besagt nun, wie der Erkenntnisprozeß vor sich gehen sollte: dies ist nichts anderes als das Bayes-Theorem. Bei jeder Wahrscheinlichkeit, die wir einer Hypothese zuweisen, handelt es sich um eine bedingte Wahrscheinlichkeit, bedingt durch die uns zur Verfügung stehende Information. Wenn sich diese verändert (vermehrt), ist die Wahrscheinlichkeit, die wir dieser Hypothese beimessen, entsprechend zu revidieren. Diese Vorgehensweise bildet die Grundlage des Lernens aus Erfahrung. Die Formalisierung dieses Prozesses geschieht über Regel 3 anhand des Bayes-Theorems: A-posteriori-Wahrscheinlichkeiten resultieren aus der Gewichtung der A-priori-Wahrscheinlichkeit mit der Evidenz der Daten anhand der Likelihoodfunktion.

²⁵⁶ Dies ist ein entscheidender Punkt: Er bedeutet, daß nicht nur dem Endergebnis, sondern auch der Vorgehensweise in einer Untersuchung Rechnung getragen werden sollte.

²⁵⁷ Jeffreys (1961, S. 11).

²⁵⁸ Ebda., S. 401. Damit ist eindeutig, daß Jeffreys *nicht*, wie des öfteren fälschlich behauptet wurde, “objektiver” Bayesianer war.

²⁵⁹ Trotzdem kann die Hypothese gemäß Regel 4 falsch sein.

Ein weiterer bedeutender Wegbereiter der modernen bayesianischen Wahrscheinlichkeitstheorie war L. J. Savage.²⁶⁰ Savage gründete seine Wahrscheinlichkeitskonzeption Ende der vierziger/Anfang der fünfziger Jahre, vor allem beeinflusst durch Milton Friedman und John von Neumann, auf eine Nutzentheorie. 1954 publizierte er sein grundlegendes Werk *The Foundations of Statistics*, in dem er versuchte, die seiner Ansicht nach eher lose verbundenen von R. A. Fisher und J. Neyman/E. S. Pearson entwickelten Techniken in einen einheitlichen Rahmen einzuordnen, der auf einer Theorie der Entscheidungsfindung unter Unsicherheit beruhen sollte.²⁶¹

Einen anderen Weg schlug H. E. Robbins (1955) ein.²⁶² Er setzte keine epistemischen, sondern "objektive" A-priori-Wahrscheinlichkeiten voraus. Er ging von der Frage aus, ob man den Bayes-Ansatz auch verwenden könne, wenn die A-priori-Wahrscheinlichkeit eines Parameters zwar unbekannt sei, aber "existiere". Die Annahme einer objektiv existierenden A-priori-Wahrscheinlichkeit wird von den meisten "Bayesianern" allerdings nicht geteilt oder – positiv gesprochen – nicht benötigt.²⁶³

Bei den bisher angesprochenen "bayesianischen" Arbeiten hatten wir den Aspekt der Wahrscheinlichkeit in den Vordergrund gestellt. Wir wollen uns nun dem zweiten Teil der bayesianischen Inferenz, dem Likelihood-Teil, zuwenden. Die nun folgenden Likelihood-Ansätze entstanden zunächst unabhängig von bayesianischen Konzepten.

Die von Fisher entwickelten Likelihood-Ideen wurden vor allem von Barnard²⁶⁴ fortentwickelt. Eine grundlegende theoretische Fundierung erhielten diese Ideen durch die wegweisende Arbeit von Birnbaum²⁶⁵, der sie zu einem Likelihood-Prin-

²⁶⁰ Leonard Jimmie Savage [1917-1971] studierte an den Universitäten Wayne und Michigan, promovierte in Mathematik mit einer Arbeit auf dem Gebiet der analytischen Geometrie und wurde 1944 nach kurzem Aufenthalt in Princeton Mitglied der *Statistical Research Group* an der Universität Columbia. 1949 gründete er mit anderen das *Statistics Department* der Universität Chicago, ging von dort 1960 nach Michigan, 1964 nach Yale, wo er zusammen mit Frank Anscombe den dortigen statistischen Fachbereich gründete. Nach Lindley (1972, S. 462) hat Savage seinen 1954 veröffentlichten Ansatz unabhängig von Ramsey und de Finetti entwickelt, später aber deren Vorgängerschaft anerkannt.

²⁶¹ Die Ausarbeitung zeigte jedoch, daß dieses Unterfangen zum Scheitern verurteilt war: "For ironically, his demonstration achieved the opposite of his intent; as it turned out almost all the techniques he had sought to justify with his coherent analysis failed to meet his test. The techniques themselves were incoherent." Lindley (1972, S. 692). Wir wollen auf die sehr technischen Ausführungen hier nicht weiter eingehen. In der zweiten Auflage des Buches äußerte sich Savage selbst zu diesem Scheitern. Siehe oben, S. 39.

²⁶² Herbert E. Robbins [1915-] studierte in Harvard Mathematik, arbeitete dort seit 1938 als Topologe am *Institute for Advanced Studies*, ein Jahr später an der Universität von New York, während des Zweiten Weltkrieges für die Marine. Sein Interesse für wahrscheinlichkeitstheoretische Fragen entstand im Zusammenhang mit einem Problem ineffizienter Bombenangriffe. 1953 kehrte er an die Universität von Columbia zurück.

²⁶³ I. J. Good z. B. zählt jedoch in der ihm eigenen Art eine ganze Batterie von unterschiedlichen "hierarchischen", "pseudo-" und "quasi-" Bayes-Ansätzen auf, in denen dieser empirische Bayes-Ansatz mittlerweile anzutreffen sei. Vgl. Good (1992, passim).

²⁶⁴ Barnard (1947), Barnard (1949). G. Barnard arbeitete nach dem Studium der Mathematik während des Zweiten Weltkrieges im Bereich industrieller Anwendungen für die englische Regierung. Zur historischen Entwicklung siehe Berger/Wolpert (1988, S. 22ff).

zip (LP) ausbaute.²⁶⁶ Zu dieser Zeit dominierte in der Statistik bereits der Neyman-Pearson-Ansatz und dessen entscheidungstheoretische Weiterentwicklung durch A. Wald.²⁶⁷

A. Birnbaum wies nach, daß das Likelihood-Prinzip zum einen aus zwei bis dahin allgemein akzeptierten Prinzipien, der Suffizienz (S) und der Konditionalität (K),²⁶⁸ folgte, zum anderen diese beiden aber auch implizierte: $(LP) \Leftrightarrow (S) \wedge (K)$. Das Likelihood-Prinzip hatte radikale Konsequenzen. Es konstatierte, daß die gesamte Evidenz eines Datums in der Likelihoodfunktion enthalten sei. Somit sei, *nachdem* die Daten erhoben wurden, der Stichprobenraum irrelevant.²⁶⁹ Das bedeutet, daß Evidenzmaße, die sich auf den Raum aller möglichen Daten beziehen (auf den Wahrscheinlichkeits- oder Parameterraum), wie etwa P -Werte oder Konfidenzniveaus, *nach* Realisation eines konkreten Datums für die Inferenz ohne Belang sind. Dies war eine Absage an die frequentistische Position, ohne auf bayesianische Argumente zurückgreifen zu müssen.

Eine Reihe von kritischen Punkten wurde in der Literatur gegen das Likelihood-Prinzip vorgebracht:²⁷⁰

- (i) Das Modell müsse vollständig parametrisiert (und richtig) sein. Die Likelihoodfunktion sei nur dann "wahr", wenn das Modell "wahr" ist. Die Likelihoodfunktion enthalte die gesamte Evidenz eines Datums *nur unter der Bedingung, daß ein Modell vorgegeben ist*. Falls man ein anderes Modell unterstelle, folge daraus eine andere Likelihoodfunktion. Dieser Vorbehalt gilt allerdings für jedes Inferenzverfahren.²⁷¹
- (ii) Es gäbe Situationen, in denen ein Kriterium, das auf Häufigkeiten basiert, von Interesse ist (z. B. Produktionskontrolle).
- (iii) Maximum-Likelihood-"Schätzungen" seien manchmal "schlecht" (in einem häufigkeitstheoretischen Sinne). Das ist aber kein Gegenargument, da das LP keine Implementation vorschreibt.

²⁶⁵ A. Birnbaum [1923-1976] studierte in Berkeley und Los Angeles Mathematik und Philosophie (unter anderem bei H. Reichenbach). 1954 promovierte er in mathematischer Statistik in Columbia, anschließend ging er zum *Courant Institute of Mathematical Sciences*, wo er 1963 Professor für Statistik wurde. Während dieser Zeit war er von 1956 bis 1957 Gastprofessor am *Imperial College*, London und 1959 Gastforscher am *Jackson Laboratory* für statistische Genetik in Bar Harbor. Nach einer Gastprofessur am *University College*, London, wurde er 1975 Professor für Statistik an der *City University* (London), wo er bis zu seinem Tode blieb.

²⁶⁶ Birnbaum (1962). Vgl. zum folgenden auch Bjornstad (1992).

²⁶⁷ Siehe oben, Abschnitt A 3 e.

²⁶⁸ Siehe dazu oben, Abschnitt A 3 d.

²⁶⁹ Siehe dazu die Beispiele in Anhang A1.

²⁷⁰ Siehe die Zusammenstellung bei Berger/Wolpert (1988) mit den ebenfalls dort zu findenden Gegenargumenten.

²⁷¹ Auf diesen Aspekt kommen wir im übernächsten Kapitel noch einmal zurück.

- (iv) Das Likelihood-Prinzip beziehe sich immer nur auf die Evidenz eines gegebenen Datums ("Experimentes"). Hierin ist jedoch nicht unbedingt ein Nachteil zu sehen.
- (v) In vielen Situationen sind die numerischen Schlußfolgerungen ähnlich oder sogar identisch, wie z. B. bezüglich von Parametern aus Normalverteilungen.²⁷² Auch dieser Punkt ist unseres Erachtens weder ein Vor- noch ein Nachteil.

Die Verwendung der Likelihoodfunktion im Rahmen einer Inferenzkonzeption wollen wir hier noch einmal in den Worten W. Stegmüllers charakterisieren:²⁷³

"1. Verwendung des Begriffs des statistischen Datums in dem [...] weiteren Sinn, wonach theoretisches Hintergrundwissen in der Gestalt akzeptierter statistischer Oberhypothesen in das Datum einzuschließen ist [d. h., daß die Verteilungsfunktion vorgegeben ist, T. R.].

2. Ausschließliche Betrachtungen von Klassen miteinander rivalisierender Hypothesen, niemals jedoch isolierter Hypothesen [allerdings in direkter Interpretation, nicht in Hinblick auf eine hypothetische Häufigkeitsverteilung auf lange Sicht, T. R.].

3. Wahl eines bloß komparativen Begriffes als Ausgangsbasis."²⁷⁴

Diese drei Eigenschaften, die die Likelihoodfunktion und die auf sie gegründete Stützungslogik im übrigen von der Carnapschen Bestätigungsfunktion und der auf diese gegründete Bestätigungslogik unterscheiden, sind noch durch eine vierte Eigenschaft zu ergänzen, die den stärksten Widerspruch zu der häufigkeitstheoretischen Position darstellt:

4. Die gesamte für die Hypothese relevante Evidenz ist in der Likelihood-Funktion enthalten; der Stichprobenraum ist somit irrelevant.

Wenden wir uns nun noch der Verbindung von A-priori-Wahrscheinlichkeiten und der Likelihood-Inferenz zur bayesianischen Inferenz zu. Der Durchbruch in praktischer Hinsicht gelang einem Bayes-Ansatz schließlich 1963 durch eine Arbeit von W. Edwards, H. Lindman und L. J. Savage, die derartige Ansätze erstmals einem breiteren Publikum zugänglich machte.²⁷⁵

Edwards, Lindman und Savage gaben zunächst einen allgemeinverständlichen Überblick über die Grundideen subjektivistischer Wahrscheinlichkeitsauffassungen und bayesianischer Analysen.²⁷⁶ Anschließend gingen sie dem zentralen

²⁷² Vgl. allerdings die Gegenbeispiele in Anhang A1.

²⁷³ Vgl. Stegmüller (1973, S. 15ff, insbesondere S. 25), Menges (1976, S. 310f).

²⁷⁴ Stegmüller fügt ergänzend hinzu, daß dieser Begriff nicht durch die Wendung 'ist wahrscheinlicher als' wiedergegeben werden darf, da er nicht die formale Struktur einer Wahrscheinlichkeit habe; die ist aber im Falle einer Standardisierung der Likelihoodfunktion sehr wohl möglich. Vgl. Menges (1976, S. 310f).

²⁷⁵ Edwards/Lindman/Savage (1963 [1992]). Wir wollen im folgenden nur auf einige Ideen, nicht auf technische Einzelheiten eingehen.

²⁷⁶ Ebda., S. 534-540.

Vorbehalt gegen bayesianische Ansätze nach, wie denn wissenschaftliche Objektivität möglich sein könne, wenn unterschiedliche Wissenschaftler unterschiedliche A-priori-Ansichten und damit A-priori-Wahrscheinlichkeit(sverteilung)en mitbringen. Sie verwandten dabei nicht das von Laplace und Edgeworth vorgebrachte Argument, daß mit wachsendem Umfang der Daten der Einfluß der A-priori-Verteilung immer geringer werde, um schließlich ganz zu verschwinden, sondern gingen der Frage nach, wann denn eine A-priori-Verteilung als Gleichverteilung angenommen werden könne bzw. die genaue Form der A-priori-Verteilung von keiner großen Bedeutung für die A-posteriori-Verteilung sei. Sie zeigten, daß "it suffices that your actual prior density change gently in the region favored by the data and not itself too strongly favor some other region."²⁷⁷ Diese vagen Bezeichnungen wurden anschließend in eine mathematische Form gefaßt und zeigten, daß eine solche Vorgehensweise in der Tat unter recht schwachen Voraussetzungen gerechtfertigt ist.²⁷⁸

Von großer Bedeutung ist für Edwards, Lindman und Savage der Aspekt des "Bayesianischen Hypothesentestens". Wenn eine Alternative zur dominierenden klassischen Statistik geboten werden sollte (und das war ihr Anspruch), dann mußte diese auch einen so zentralen Aspekt wie das Testen wissenschaftlicher Hypothesen beinhalten.²⁷⁹ Zunächst klärten sie die Begriffe "odds" und "likelihood ratios". Anhand des Beispiels einer Überprüfung, ob ein Würfel "fair" ist, wurde dann die Verwendung des Likelihood-Quotienten in bayesianischer Sichtweise der klassischen Verwendung (nach Neyman/Pearson, s. o.) gegenübergestellt. Insbesondere die Problematik der in der klassischen Statistik favorisierten Betrachtung von Fehlern erster und zweiter Art anhand dieser Testgröße wurde von ihnen verdeutlicht:

"The interesting point is made that a Bayesian hypothesis test can add extensive support to the null hypothesis whenever the likelihood ratio is large. The classical test can only reject hypotheses, and it is not clear just what sort of evidence classical statistics would regard as a strong confirmation of a null hypothesis."²⁸⁰

Wir wollen auf die – zumeist sehr technischen – Einzelheiten an dieser Stelle verzichten. Mittlerweile ist für zahlreiche Einzelprobleme und auch für grundsätzliche Fragen, wie die bayesianische Interpretation häufigkeitstheoretischer Standpunkte,²⁸¹ rein empirische bayesianische Ansätze²⁸² oder gar eine Theorie der bayesianischen Datenanalyse, eine Lösung angeboten worden.²⁸³

²⁷⁷ Ebda., S. 541. Dies wird als "stable estimation" bezeichnet.

²⁷⁸ DuMouchel (1992, S. 521) weist darauf hin, daß dieser Ansatz in enger Beziehung zu den später von anderen Bayesianern vorgeschlagenen 'reference priors' steht, die in Situationen angewendet werden können, in denen nur wenig A-priori-Information vorhanden ist und die auch für klassische Statistiker akzeptabel sind. Auf der anderen Seite, so räumten die Autoren ein, gäbe es aber auch Situationen, in denen die genauen Eigenschaften der A-priori-Verteilung entscheidend seien. Sie zählten fünf solcher Situationen auf. Siehe Edwards/Lindman/Savage (1963 [1992, S. 546]).

²⁷⁹ Die Notwendigkeit einer Testtheorie wird in der bayesianischen Literatur nicht einheitlich gesehen.

²⁸⁰ DuMouchel (1992, S. 523). Siehe dazu Beispiel 3 in Anhang A2.

Eine wichtige Frage ist in diesem Zusammenhang die Beurteilung von Signifikanztests und Konfidenzintervallen.²⁸⁴ Die Verwendung von Signifikanztests in ihrer häufigkeitstheoretischen Position wird von einer Reihe von Bayesianern als heuristisches Hilfsmittel durchaus befürwortet.²⁸⁵ Falls die A-priori-Information gering ist, können die Konfidenzintervalle der klassischen Statistik und bayessche Wahrscheinlichkeitsintervalle numerisch nahezu identisch sein. Sie sind aber völlig unterschiedlich zu interpretieren.²⁸⁶ Nach der klassischen, frequentistischen Interpretation bedeutet ein 95 %-Konfidenzintervall, daß bei gegebenen (gleichen) Stichprobenumfängen n für $m \rightarrow \infty$ (m = Anzahl der Stichproben) 95 % der Intervalle den wahren, unbekannten, festen Parameter abdecken und 5 % nicht. Wir wissen dagegen nicht (können es nur hoffen), ob das konkret vorliegende Intervall den Parameter abdeckt oder nicht. Die bayessche Analyse unterstellt dagegen dem unbekannten Parameter eine (in der Regel subjektive) A-priori-Verteilung. Nachdem Daten erhoben wurden besteht weiterhin eine Unsicherheit, nun aber eine geringere. Diese Unsicherheit wird zwar auch in Wahrscheinlichkeiten ausgedrückt, aber mit einer völlig anderen Interpretation: der Parameter θ liegt mit 95 %iger Wahrscheinlichkeit zwischen zwei Werten c_u und c_o . Solch eine Interpretation ist im Sinne der klassischen Inferenzstatistik nicht möglich,²⁸⁷ obwohl sich bis heute mißverständliche Interpretationen dieser Bayes-Epistemologie in der klassischen Literatur finden.²⁸⁸

²⁸¹ Vgl. Pratt (1965).

²⁸² Siehe Morris (1983).

²⁸³ Hill (1990). Auf diesem Gebiet kann der bayesianische Standpunkt in besonderem Maße Vorteile verbuchen. Siehe jetzt auch Gelman/Carlin/Stern/Rubin (1995).

²⁸⁴ Siehe hierzu grundsätzlich Hodges (1990).

²⁸⁵ Siehe dazu ebda. Anderer Meinung sind etwa Jeffreys (1961, S. 383-398), Berger/Wolpert (1988) oder Lindley in Box (1980, S. 423). Wie wir gesehen haben, läßt sich die ambivalente Haltung in dieser Frage bis zu Edgeworth und Laplace zurückverfolgen. Siehe oben, Abschnitte A 3 b, c.

²⁸⁶ Das folgende nach Iversen (1984, S. 31).

²⁸⁷ Ebda.: "This is the way many users of confidence intervals want to interpret a confidence interval, but in classical statistical inference such an interpretation is not possible."

²⁸⁸ Eine häufig anzutreffende, mißverständliche Interpretation mag hier mittels eines einzigen Beispiels illustriert werden. Wonnacott/Wonnacott (1990) betrachten in ihrem verbreiteten Lehrbuch den bayesschen Ansatz als Erweiterung der klassischen Statistik, in dem zusätzlich zu den dort verwendeten Verfahren noch weitere (A-priori-) Informationen Verwendung finden. Die klassische Statistik wird von ihnen in den ersten 18 (von 20), die Bayes-Statistik daran anknüpfend in den letzten beiden Kapiteln behandelt. Die unterschiedliche Bedeutung von Konfidenzintervallen im klassischen und bayesschen Kontext wird in diesem Zusammenhang nicht erläutert. Die Autoren erwähnen lediglich (S. 615), daß die (A-priori-) Zusatzinformation zu engeren Konfidenzintervallen führt, mit zunehmendem Stichprobenumfang diese A-priori-Information aber immer unbedeutender werde. Im Grenzfall sind für Wonnacott/Wonnacott sogar beide Ansätze identisch: "We must remember that as the sample becomes large enough (and hence swamps the prior) such posterior information reduces to the classical equivalent. For example, the posterior confidence interval [eine zumindest ungewöhnliche Bezeichnung, T. R.] reduces to the classical confidence interval; and the whole posterior distribution reduces to the classical likelihood function. (This is yet another reason to study classical statistics for 18 chapters.)" Es ist nochmals zu betonen, daß diese numerische Übereinstimmung eben *keine* inhaltliche Übereinstimmung bedeutet.

Wir müssen jedoch festhalten, daß es *den* bayesianischen Ansatz nicht gibt, sondern ein Konglomerat unterschiedlicher Ansätze vorliegt, denen zumindest ihre Ablehnung der “klassischen” Sichtweise gemein ist, auch wenn man sich in den letzten Jahren vielleicht wieder näher gekommen ist:

“In the last ten years or so the polarization between supporters and opponents of Bayesian methods has decreased, but one still often hears a statistician say ‘I am not a Bayesian.’ The meaning is often unclear because of the 6⁶ meanings of ‘Bayesian.’ In my opinion all statisticians are Bayesian, or at least use a Bayes/non-Bayes compromise, in any applied problem. Therefore, when a statistician says ‘I am not a Bayesian,’ it seems to me somewhat like saying ‘I am very modest.’ Another analogy would be a psychologist saying ‘The result of this experiment, in which experimental dissimulation was used, can be trusted because I never tell a lie.’”²⁸⁹

4. Aspekte der Modellbildung

a. Allgemeines

Die bisherigen Ausführungen setzten implizit voraus, daß im Moment der Inferenz/Parameterschätzung das Modell bereits gegeben ist. Bereits im einfachsten Falle einer Schätzung eines Anteilswertes wird von einem Modell ausgegangen: Die postulierten Wahrscheinlichkeitsgesetze gelten nur, wenn die Beobachtungen unabhängig voneinander sind.

Stegmüller (1973, S. 240ff) spricht im Zusammenhang von Tests von “Oberhypothesen” bzw. “Hintergrundwissen”, die als Voraussetzung nicht hinterfragt werden oder zumindest nicht Gegenstand eines Tests sind. Er führt ein illustratives Beispiel an: Er geht von einer experimentellen Situation aus, in der die Wahrscheinlichkeiten der Augenzahlen für Würfe mit einem bestimmten Würfel ermittelt werden sollen. Bei der Durchführung dieses Experiments gehe man – bewußt oder unbewußt – von Oberhypothesen aus, etwa, daß die Würfe unabhängig seien oder eine bestimmte Verteilung vorliege. Nach einer Weile stelle man etwas *Unerwartetes* oder “Seltsames” fest: Beim Werfen treten niemals Dreier-Folgen auf. Immer, wenn eine Augenzahl dreimal hintereinander gewürfelt wird, geschieht dies auch ein viertes Mal. Stegmüller (1973, S. 240f) fragt nun: Was aber heißt es, daß diese Ergebnisse *seltsam* sind? Sie sind nur seltsam *unter der Annahme, daß Unabhängigkeit vorliegt*.²⁹⁰ Und weiter: “Es kann z. B. ein Konflikt bestehen zwischen dem, was die Daten lehren, wenn sie testtheoretisch ausgewertet werden, und den Einfachheitsüberlegungen, die für die Annahme der ursprünglichen Oberhypothese sprechen.” Zwei Eindrücke stehen sich nun gegenüber: auf der einen Seite die “Macht der Einfachheit” (Unabhängigkeitsannahme), auf der anderen Seite die “Schwachheit der Daten”. Die entscheidende Frage sei, wann der Eindruck des letzteren zu der Überzeugung führt, daß zumindest eine Oberhypothese falsch ist. Stegmüller konstatiert: “Die theoretische Statistik muß vorläufig zugestehen, daß

²⁸⁹ Good (1983a, S. 95). Vgl. auch Good (1988, S. 295).

²⁹⁰ Hervorhebungen im Original.

sie für derartige Konfliktsituationen überhaupt über keine präzisen Kriterien verfügt.” (S. 241).²⁹¹

Die Relevanz dieser Frage für unser Thema wird besonders deutlich, wenn Metz (1995, S. 466) nach einer umfangreichen vergleichenden Zusammenstellung verschiedener Test- und Schätzverfahren für stochastische Trends in univariaten Zeitreihen/ Prozessen am Ende feststellt, “[...] daß die hier abgeleiteten Ergebnisse bedingte Ergebnisse darstellen: Sie gelten nur unter der Bedingung der Gültigkeit des gewählten Modells.”

Überblickt man die Literatur zur Beziehung von Inferenz und Modellbildung,²⁹² so fällt auf, daß die Frage nach der Rolle von Modellen in der Regel erst dann thematisiert wird, *nachdem* ein expliziter Standpunkt hinsichtlich der Inferenzkonzeption eingenommen wurde. Eine erste Orientierung zum Thema bieten wiederum die entsprechenden Artikel der *Encyclopedia of the Statistical Sciences*.²⁹³ Koch/Gillings (1983) unterscheiden drei Situationen, in denen (statistische) Modelle Verwendung finden:

1. Experimentelle Studien von zufällig ausgewählten Subjekten,
2. historische oder beobachtende Studien aller Subjekte einer zufälligen oder natürlichen Population und
3. Stichprobenerhebungen zufällig ausgewählter Subjekte.

Generell sei zwischen design- und modellbasierter Inferenz zu unterscheiden:

“A statistical analysis whose only assumptions are random selection of observational units or random allocation of units to experimental conditions may be said to generate *design-based inferences* [...]. Also, such inferences are often said to have *internal validity* [...] when the design is adequate to eliminate alternative explanations for the observed effects other than the one of

²⁹¹ Einige instructive Gedanken zu dieser Problematik finden sich bereits bei E. Claparède (1932 [1969]), die jedoch in der statistischen und erkenntnistheoretischen Literatur, soweit wir übersehen, nicht rezipiert wurden.

²⁹² Untersuchungen, die Auswirkungen der Modellselektion auf Parameterschätzungen thematisieren, beschränken sich in der Regel auf “automatische” Modellselektionsprozeduren, die von einer gegebenen Klasse von Modellen ausgehen, wobei die Auswirkungen dann relativ unproblematisch formal untersucht werden können, so etwa Pötscher (1991). Zwar sind etwa die sog. “Informationskriterien”, wie das Akaike Informationskriterium (AIC) oder das bayesianische Informationskriterium (BIC) – das im übrigen oftmals im Rahmen klassischer Studien verwendet und dessen Qualität aus häufigkeitstheoretischer Warte, d. h. auf lange Sicht, untersucht wird (vgl. etwa Schlittgen/Streitberg (1994, S. 339ff)) – zum Teil auch theoretisch fundiert (de Leeuw (1992, S. 600)): “[...] Akaike and his colleagues succeeded in connecting the AIC effectively to the always mysterious area of the foundations of statistics”) und gehen somit über heuristische Spezifikationsverfahren hinaus, ihre Anwendung bedeutet aber lediglich eine Minimierung bestimmter Kriterien unter der Bedingung, daß die in Frage kommenden Modelle vorgegeben sind.

²⁹³ N.N. (1985) enthält lediglich kurze terminologische Definitionen. Der offensichtlichen Unterrepräsentation des Modellaspektes wurde durch einen weiteren Artikel im Supplementband Rechnung getragen. Die Literatur zu diesem Aspekt steht in keinem Verhältnis zu der ausufernden Literatur bezüglich klassischer versus bayesianischer Inferenz oder schätztechnischer Spezialprobleme.

interest. In this sense, internal validity requires only that the *sampled population* and the *target population* be the same.”²⁹⁴

Das ist zweifellos die Situation, die R. A. Fisher vor Augen hatte.

“Alternatively, if assumptions external to the study design are required to extend inferences to the target population, then statistical analyses based on postulated probability distributional forms (e.g. binomial, normal, Poisson, Weibull etc.) or other stochastic processes yield *model-based inferences*. These can be viewed as encompassing *Bayesian inferences* and *superpopulation inferences* to the extent that the validity of the claimed generality is model dependent via its sensitivity to model misspecifications.”²⁹⁵

Letztere Definition ist somit eine umfassendere. Auf der anderen Seite kann aber die Frage des Ausmaßes der Repräsentativität nicht beantwortet werden: “their principal limitation is that subjects in a study may not represent any meaningful population beyond themselves.”²⁹⁶

Auf jeden Fall sollte man sich der folgenden Tatsache bewußt sein: Ein reines Mustererkennen ist nicht sinnvoll, eine wie auch immer geartete “Theorie” muß stets vorhanden sein. Zwar hat es in der Wissenschaft immer wieder postulierte Gesetze gegeben, die aus subjektiv wahrgenommenen Regelmäßigkeiten geboren wurden, doch haben sich solche Gesetze in der Regel als unhaltbar erwiesen.²⁹⁷

Nun gibt es zahlreiche Umstände, in denen nicht immerwährende Gesetze, sondern solche, die sich im Zeitablauf ändern, angenommen werden können. In solchen Situationen ist es nicht immer sinnvoll, a priori von Theorien auszugehen und diese lediglich zu überprüfen. Geht man hier jedoch rein explorativ vor, kann man in zweifacher Hinsicht scheitern (“Fehler erster und zweiter Art begehen”). Einerseits ist es möglich, daß man existierende Zusammenhänge/Strukturen in den Daten nicht findet, weil sie durch die “Explorierungsverfahren” nicht aufgedeckt werden können, andererseits können Zusammenhänge/Strukturen “gefunden” werden, die aber nur für diese Daten zutreffen, also “zufällig” sind, aber nicht auf eine zugrundeliegende Struktur schließen lassen. Dabei muß es sich nicht unbedingt um eine “zufällig” in den Daten vorhandene Struktur handeln (also um eine stochastische Konzeption), es kann auch eine falsche Anpassung sein. Insofern sind

²⁹⁴ Koch/Gillings (1983, S. 85). Hervorhebungen im Original.

²⁹⁵ Ebda. Hervorhebungen im Original.

²⁹⁶ Ebda., S. 86.

²⁹⁷ Ein bekanntes Beispiel ist das sog. Bodesche Gesetz von 1772. Demnach ist die relative Entfernung der Planeten von der Sonne durch das folgende “Gesetz” gekennzeichnet: Normiert man die Entfernung der Erde zur Sonne auf 1, so ergibt sich die Entfernung aller anderen Planeten wie folgt: Zu der Zahl 4 werden sukzessive die Zahlen 0, 3, 6, 12, 24, 48, 96 usw. addiert und anschließend durch 10 dividiert. Die so ermittelten Zahlen geben die relativen Entfernungen der Planeten zur Sonne wieder. Dieses “Gesetz” traf nicht nur für alle bis dahin bekannten Planeten zu, sondern bewies auch noch seine Gültigkeit, als 1781 der Uranus entdeckt wurde. Selbst die letzten Skeptiker mußten sich schließlich von der universalen Gültigkeit dieses Gesetzes überzeugen lassen, als man feststellte, daß alle relativen Entfernungen dieser Reihe bis auf 2.8 besetzt waren und man daraufhin in dieser Lücke tatsächlich den Asteroiden Ceres entdeckte. Doch bereits die Entdeckung des Neptun an Position 30 anstelle der gesetzmäßigen Position 39 und des Pluto bei 40 anstelle von 77 machten das ermittelte “Gesetz” hinfällig. Vgl. Davis (1941, S. 4).

auch explorative Verfahren nicht “frei”. Diese Probleme stellen sich, *bevor* überhaupt Parameter eines Modells geschätzt werden.

Wir wollen im folgenden auf zwei Aspekte eingehen, die für unseren Zusammenhang von besonderer Bedeutung sind und untersuchen, wie diese Probleme jeweils beantwortet wurden. Es handelt sich hierbei um die Rolle der Modellbildung in der Zeitreihenanalyse und die Bedeutung der sog. “explorativen” Datenanalyse, speziell den zumeist mit dieser Bezeichnung in Verbindung gebrachten Ansatz von John W. Tukey und – als Vergleich dazu – den explizit bayesianische Elemente enthaltenen Ansatz von George E. P. Box. Zunächst jedoch zur Zeitreihenanalyse.

b. Der Faktor Zeit: ‘Die ganze Welt ist ein stochastischer Prozeß’

Die allgemeinen Bemerkungen im vorigen Abschnitt hatten einen Aspekt ausgeklammert, der nicht nur für unser Thema von zentraler Bedeutung ist: die Dimension der Zeit. Wie Menges zu Recht feststellt, sind “die meisten Erscheinungen der Natur und der Gesellschaft [...] keine Einzelereignisse, auch keine zusammengesetzten stochastischen Ereignisse, auch keine Zufallsvariablen, sondern zufallsabhängige *Vorgänge* [...]”. Der Begriff des stochastischen Prozesses ist so außerordentlich gemein, daß man sich fragen muß, welches in der Zeit veränderliche Sein ihm nicht unterworfen ist. Die ganze Welt läßt sich als ein stochastischer Prozeß auffassen.²⁹⁸

In der historischen Entwicklung der Zeitreihenanalyse sind zwei Linien zu trennen: zum einen die *Analyse konkreter Zeitreihen*, die in den verschiedensten Zusammenhängen beobachtet wurden, zum anderen die Theorie *stochastischer Prozesse*. Beide finden durch die Arbeiten von G. U. Yule und E. Slutsky etwa um 1927 eine erste Symbiose, weisen zu dieser Zeit aber bereits längere eigenständige Traditionen auf.

1. Die *Analyse konkreter Zeitreihen* widmet sich vor allem folgenden Problemen:

- der Bestimmung von “Trends”,
- der Untersuchung von Zyklen
- und der Übertragung der Korrelationskonzeption von Querschnitt- auf Längsschnittdaten.

Die Wurzeln liegen auch hier in der Astronomie. Die Beobachtungen der Planetenbahnen, in deren Rahmen diese Fehlertheorie entstand (s. o.), waren nichts

²⁹⁸ Menges (1972, S. 145, 146). Hervorhebung im Original. Stegmüller widmet dem Problem dagegen keine große Aufmerksamkeit. Falls sich Wahrscheinlichkeiten im Zeitablauf ändern, müsse man von “Momentphotographien” ausgehen. Für diese könne man dann wieder konstante Gesetze unterstellen. Vgl. Stegmüller (1973, S. 35). Auch R. A. Fisher hat dem Problem wenig Aufmerksamkeit gewidmet. Dies können wir einer Äußerung von G. E. P. Box entnehmen: “I remember that when I started getting interested in time series I told him, and he sounded rather disappointed and said, ‘Oh I don’t think there’s much in that.’ He wasn’t all that enthusiastic about time series.” Box in DeGroot (1987, S. 254).

anderes als Zeitreihen. In diesem Zusammenhang stellte man fest, daß sich im Laufe der Zeit neben den periodischen auch langsame, scheinbar nichtperiodische (“säkulare”) Veränderungen vollzogen.²⁹⁹ Die Gesetzmäßigkeiten für diese Veränderungen versuchten die größten Mathematiker des 18. und 19. Jahrhunderts zu ergründen, darunter Laplace, Euler und Lagrange. 1787 konnte Laplace nachweisen, daß die langfristigen Veränderungen der Bahnen des Jupiter und Saturn einer Periode von 900 (!) Jahren unterlagen. Auch für die “säkulare” Bewegung des Mondes konnte eine solche lange Periode nachgewiesen werden. Damit waren weitere Argumente für den mechanischen Aufbau der Welt gefunden. Diese “Komponentenmodelle” wurden von der Astronomie wiederum in andere Wissenschaftsbereiche übernommen.³⁰⁰ Einige Beispiele illustrieren die schnelle Verbreitung dieses Ansatzes:³⁰¹

1801	Herschel	Untersuchung der Beziehung zwischen Sonnenflecken und Weizenpreis
1832	Forbes	Beginn einer Reihe von meteorologischen/barometrischen Studien
1843	Guy	Krankheit/Todesfälle in Abhängigkeit der Saison/Temperatur
1843	Everest	Hungerkrisen in Indien in Abhängigkeit vom Weizenpreis
1844	Verhulst	erste (?) Berechnung einer logistischen Trendfunktion ³⁰²
1852	Hain	Trendberechnung nach der Kleinstquadratmethode ³⁰³

Für eine Reihe von empirischen Erscheinungen verfestigte sich die Überzeugung, daß sie sich aus verschiedenen Komponenten zusammensetzten, vor allem aus voneinander unabhängigen zyklischen Komponenten unterschiedlicher Dauer, wobei oftmals eine “saisonale” Komponente die offensichtlichste war. Vor allem durch C. Babbage und W. S. Jevons fanden diese Überzeugungen Eingang in die Ökonomie.³⁰⁴ Jevons “entdeckte” in zahlreichen Verläufen einen Zyklus von etwa elfjähriger Dauer, für den er die ebenfalls mit einer Periode von elf Jahren zu- und abnehmende Anzahl von Sonnenflecken als letzten Grund verantwortlich machte. Er untersuchte möglichst lange agrarwirtschaftliche Reihen sowie Kreditreihen des 19. Jahrhunderts.

Konstituierend für die angesprochenen Arbeiten war, daß sie nicht von theoretischen Überlegungen, sondern von beobachteten und angenommenen empirischen

²⁹⁹ Vgl. Nerlove/Grether/Carvalho (1979, S. 2ff).

³⁰⁰ Nach ebda., S. 2, fand dieser Transfer zwischen 1825 und 1875 statt.

³⁰¹ Siehe, wenn nicht anders angegeben, ebda., S. 4f, dort auch die entsprechenden Nachweise. Die meisten Arbeiten wurden hier von der Meteorologie beeinflusst.

³⁰² Vgl. Davis (1941, S. 18).

³⁰³ Vgl. Stigler (1986, S. 312). Dies dürfte wohl die früheste Anwendung der Methode sein; sie fand jedoch keine weitere Beachtung außer einer Notiz von Edgeworth.

³⁰⁴ Nerlove/Grether/Carvalho (1979, S. 354-360) behandeln auch die Arbeiten von B. Ballot, die jedoch keinen Einfluß auf die weitere Entwicklung hatten.

Regularitäten ausgingen. Hier standen keine inferenzstatistischen Aussagen im Vordergrund, sondern eher deskriptive Anpassungen.³⁰⁵ In methodischer Hinsicht waren diese Arbeiten durch Ad-hoc-Überlegungen gekennzeichnet. Jevons verwendete ein 11-Jahres-Gitter, das er über graphische Darstellungen der Reihen legte sowie gleitende Durchschnitte und Monats- bzw. Quartalsdurchschnitte, mit denen er die Reihen bereinigte.³⁰⁶ Zu dieser Zeit gehörte die Anwendung gleitender Mittelwerte zur Glättung von Reihen bereits zum Repertoire statistischer Untersuchungen.³⁰⁷

Insbesondere die Untersuchung von Zyklen fand um die Jahrhundertwende ein ebenso mächtiges wie problematisches Instrument: die sogenannte Periodogramm-analyse. Sie begann Ende des 19. Jahrhunderts im Rahmen harmonischer Analysen geophysikalischer Phänomene.³⁰⁸ Sie konnte sich hierbei auf Resultate der Mathematik stützen: J. L. Lagrange entwickelte 1772 erste Grundgedanken einer "harmonischen Analyse"; L. Euler hatte entdeckt, daß eine analytische Funktion als Summe trigonometrischer Funktionen ausgedrückt werden konnte; kurze Zeit später wurden diese Erkenntnisse in einer berühmten Arbeit von J. B. J. Fourier ausgebaut und systematisiert.³⁰⁹ Anwendungen dieser mathematischen Erkenntnisse (und Algorithmen) erfolgten im Zusammenhang mit statistischen Daten zunächst in der Geodäsie. Die Koeffizienten wurden hier bereits anhand der Methode der kleinsten Quadrate bestimmt.³¹⁰ Weite Verbreitung erlangten derartige Ansätze durch die Einführung einer graphischen Darstellung, dem Periodogramm 1898 durch Schuster, das im folgenden in mehrfacher Hinsicht angewendet und ausgebaut wurde. All diesen Ansätzen war die Grundüberzeugung gemein, daß das einer konkreten Reihe zugrundeliegende Phänomen ein Kompositum von Zyklen verschiedener Längen darstellte oder zumindest eine Periode aufwies, deren Signifikanz man anhand des Periodogramms testen konnte. In der zweiten Dekade des 20. Jahrhunderts fanden diese Verfahren auch Anwendung im ökonomischen Kontext (siehe unten, Kap. B 2 c).

³⁰⁵ Jevons äußerte sich in seiner 1874 erschienenen Abhandlung *The Principles of Science* hinsichtlich wahrscheinlichkeitstheoretischer Aussagen ganz in der Tradition von Laplace, Quetelet und DeMorgan und betrachtete den Zufall subjektiv als Zustand des Unwissens. Dennoch lehnte er die von Quetelet und Buckle propagierte Möglichkeit der Anwendung der Wahrscheinlichkeitsrechnung auf gesellschaftliche Phänomene ab. Wie J. Venn sah er die historische Entwicklung als zu instabil an, als daß hier konstante Gesetzmäßigkeiten unterstellt werden könnten. Vgl. Porter (1986, S. 176).

³⁰⁶ Noch rudimentärer waren die wenige Jahre zuvor von C. Juglar angewandten Methoden zur Untersuchung der in der Vergangenheit immer wiederkehrenden Handelskrisen. Juglar vermutete zwar eine Regelmäßigkeit hinsichtlich der Abfolge von Ereignissen in Krisen, lehnte jedoch die Vorstellung einer gleichen *Zeitdauer* dieser Ereignisketten und damit von Zyklen gleicher Länge strikt ab. Das hinderte Schumpeter (1939) nicht daran, im Zusammenhang von Zyklen mit einer Länge von etwa 9 bis 10 Jahren von "Juglar-Zyklen" zu sprechen. Vgl. zu Jevons Morgan (1990, S. 18-26), zu Juglar ebda., S. 41-44.

³⁰⁷ Siehe z. B. Poynting (1884).

³⁰⁸ Vgl. Brillinger (1976), Heiler (1991) und Davis (1941).

³⁰⁹ J. B. J. Fourier, *Théorie analytique de la chaleur*, Paris 1822, zitiert nach Davis (1941, S. 29).

³¹⁰ Zu den Arbeiten siehe ebda., S. 30f. Heiler (1991, S. 2) erwähnt eine bereits 1879 erschienene Arbeit "Notes on searching for periodicities" von G. C. Stokes.

Neben der Untersuchung von Trends und Zyklen wurde etwa seit der Jahrhundertwende der Versuch in Angriff genommen, die *Korrelationskonzeption* auf Zeitreihen zu übertragen. Noch lange nach Jevons Ad-hoc-Untersuchungen war es allgemein üblich, Zusammenhänge zwischen Variablen aufgrund von visuellen Inspektionen der jeweiligen Zeitreihendarstellungen zu postulieren. In vielen Fällen konnte jedoch mit gleichem Recht eine gegenteilige Interpretation herausgelesen werden.³¹¹ Insofern lag es nahe, Beziehungen durch (Korrelations-) Maßzahlen eindeutiger auszudrücken. Auch hier sollte sich das Problem des zeitlichen Horizontes wie zuvor bei der Analyse von Komponentenmodellen als zentral erweisen.

Bereits 1901 und 1905 übertrug R. H. Hooker die von Yule auf sozioökonomische Phänomene angewandte und dort weiterentwickelte Korrelationsanalyse auf Zeitreihen und unterschied hier kurz- und langfristige Zusammenhänge. Während zum Nachweis der letzteren eine einfache Korrelation der Originaldaten hinreichend sei, müsse für den Nachweis der ersteren zuvor eine Transformation vorgenommen werden, die den langfristigen Zusammenhang ausschalte. Anhand von Zeitreihen des Handels und der Heiratsrate sei folgendes festzustellen: Es sei allgemein bekannt, daß die Heiratsrate in Abhängigkeit von dem allgemeinen Wohlstand (ausgedrückt in Import-/Exportdaten) eines Landes fluktuieren. Wenn man jedoch ein Diagramm beider Reihen zeichne, sei sofort zu sehen, daß zwar die Fluktuationen korrespondieren, nicht jedoch die Reihen insgesamt:

“[...] the correspondence is clearly not absolute, because the curves further show that it is in fact only the oscillations which are correlated; and that, over the last forty years as a whole, the trade curve has arisen, whereas the marriage-rate has fallen. These latter movements may with safety be ascribed to outside causes. Now correlate the two curves (marriage-rate with, say, the value per head of the exports of British and Irish produce) for the thirty five years 1861-95, we find $r = + 0.18$ only, with a probable error of 0.09. This is a very low coefficient, indicating, no connection between the two. And the fact is that in this case the difference in the *general* movements of the two curves has completely overshadowed the minor oscillations; whereas it is only to these latter that we refer when, in ordinary parlance, we speak of the marriage-rate as being dependent upon trade.”³¹²

Weder der langfristige noch der kurzfristige, sondern der mittelfristige Zusammenhang stand also im Mittelpunkt des Interesses. Um hierfür einen Korrelationskoeffizienten ausrechnen zu können, mußte der langfristige Verlauf vorher ausgeschaltet werden. Dies sollte entweder durch die Betrachtung der Abweichungen von einem gleitenden Mittelwert oder – wenn keine regulären periodischen Fluktuationen zu erkennen seien – durch die Betrachtung von Differenzen geschehen.³¹³ Hooker vermutete weiterhin, daß die Zusammenhänge stärker sein könnten, wenn man eine zeitliche Verschiebung zwischen den Reihen berücksichtigt und präsen-

³¹¹ Vgl. Hendry/Morgan (1995b, S. 11).

³¹² Hooker (1901, S. 485f). Hervorhebung im Original.

³¹³ Vgl. auch Aldrich (1995, S. 371) mit den entsprechenden Belegen. Die Korrelation der Originaldaten sei aber dennoch sinnvoll. Sie könne Aufschluß über “‘secular’ independence” geben. Hooker (1905, S. 703).

tierte eine Berechnung für verschiedene “leads” und “lags” inklusive deren graphische Darstellung (nach heutiger Diktion: Kreuzkorrelogramme). Für die von ihm untersuchte Beziehung zwischen der Heiratsrate und verschiedenen ökonomischen Größen stellte er darüber hinaus fest, daß die Beziehungen zeitlich nicht stabil waren. Berechnungen für die gesamte untersuchte Periode von 1861 bis 1895 erbrachten andere Ergebnisse als getrennte Berechnungen jeweils für die ersten 15 und letzten 20 Jahre.³¹⁴

Hookers Analysen wurden kurz darauf von Yule fortgeführt. Yule untersuchte 1906 die Gleichläufigkeit von Heirats- und Geburtsziffern, deren gemeinsame Schwankungen gemeinhin einem Konjunkturzyklus (“trade cycle”) zugesprochen wurden und wandte dabei erstmals Verfahren an, die er wenige Jahre zuvor in Querschnittuntersuchungen vorgestellt hatte: vor allem die *partielle* Korrelation.³¹⁵ Er berechnete Korrelationen zwischen der Fruchtbarkeits- und Heiratsrate auf der einen und den Exporten, Importen, dem Handelsvolumen, dem Weizenpreis und Bank-Clearings für das gleiche und verschiedene vorangegangene Jahre auf der anderen Seite. Dabei gelangte er unter anderem zu dem Schluß, daß die Fruchtbarkeitsrate unabhängig von der Heiratsrate mit dem Handelsvolumen korrelierte. Zur Ausschaltung des Trends bediente er sich der gleichen Methoden wie zuvor Hooker. Während bei Yule und Hooker die Korrelation der “Oszillationen” im Mittelpunkt des Interesses stand, sahen andere den Sinn der Zusammenhangsanalysen von Zeitreihen vorrangig in der Berechnung kurzfristiger Abhängigkeiten. Hookers (vermeintliche) Idee von 1905 wurde auf Initiative K. Pearsons von W. Gosset (“Student”) 1914 weiter “ausgebaut”.³¹⁶ Dieser untersuchte die Beziehung zwischen der weiblichen durch Krebs verursachten Todesrate und dem Pro-Kopf-Konsum importierter Äpfel mit einer Methode der “variäten Differenzen”. Dabei ging er davon aus, daß beide Größen additiv aus einem polynomialen Trend und einer Restgröße zusammengesetzt waren. Das Interesse sollte sich auf die Restgrößen und deren Korrelation richten. Dazu mußte zuvor die polynomiale Trendkomponente (deren Potenz) anhand von Differenzen ermittelt werden.³¹⁷

Yule erhob 1921 gegen diese Generalisierung Einspruch und betonte, daß eine solche Vorgehensweise nicht der Intention Hookers entspräche:

³¹⁴ Bei der Korrelation zwischen Weizenpreis und Heiratsrate drehte sich sogar das Vorzeichen um. Für die positive Korrelation zwischen diesen beiden Größen von 1876 bis 1895 vermutete er als Grund die außergewöhnlich hohen Preise zwischen 1876 und 1881 (gewissermaßen als Ausreißer).

³¹⁵ Yule (1906).

³¹⁶ Student (1914–1915). “Student” (W. S. Gosset) war Chemiker bei der Guinness-Brauerei und ist vor allem aufgrund seiner Ableitung der 1908 von ihm unter dem Pseudonym “Student” vorgestellten *t*-Verteilung bekannt. Das von Cave/Pearson (1914–1915) kurz darauf als “variate difference Methode” bezeichnete Verfahren wurde zur gleichen Zeit unabhängig von O. Anderson entwickelt. Vgl. Anderson (1914–1915). Anderson sah diese Arbeiten als Weiterentwicklung des Ansatzes von Hooker und Cave. Mitarbeiter von Pearson hatten in dessen Laboratorium erstmals differenzierte Zeitreihen korreliert. Vgl. Cave-Brown-Cave (1904).

³¹⁷ Auf die technischen Einzelheiten wollen wir hier verzichten.

“[T]he problem is not to isolate random residuals but oscillations of different durations, and unless the generalized method can be given some meaning in terms of oscillations it is not easy to see what purpose it can serve.”³¹⁸

Hier spricht der Pragmatiker Yule also von der “Bedeutung”, die eine Analyse haben müsse und sieht diese zweifelsohne in der Bestimmung von “Oszillationen”, nicht aber von Trendpolynomen gegeben. Das Denken in (zyklischen oder zumindest glatten) Komponenten stand völlig im Vordergrund; sie galt es zu bestimmen, die Restkomponente war dagegen unbedeutend, eben zufällig (entstanden etwa durch Rundungsfehler, Beobachtungsfehler etc.), ohne Struktur oder Information. Während Yule in diesem Zusammenhang an Korrelationen zwischen (oszillatorischen) Bewegungen interessiert war und diese entweder mit einer kausalen Abhängigkeit einer Reihe von der anderen begründete oder mit einer Abhängigkeit beider von der Zeit im Sinne einer oder mehrerer Drittvariablen, von denen beide gemeinsam abhängig waren,³¹⁹ schlug er fünf Jahre später eine weitere Interpretation vor: die von ihm so bezeichnete “Nonsens-Korrelation”.³²⁰ Der Interpretation einer Bedingtheit der Korrelation durch den Faktor Zeit konnte er keinen Sinn beimessen:

“I cannot regard time per se as a causal factor; and the words only suggest that there is some third quantity varying with the time to which the changes in both the observed variables are due [...]. But what one feels about such a correlation is, not that it must be interpreted in terms of some very indirect catena of causation, but that it has no meaning at all; that in non-technical terms it is simply a fluke, and if we had or could have experience of the two variables over a much longer period of time we could not find any appreciable correlation of them.”³²¹

Aufschluß über die Beeinflussung der bivariaten Korrelation sollte statt dessen die *interne Struktur* einer Reihe geben. Als Beispiel für ein solches Phänomen führte er die Entwicklung des Anteils der Heiraten, die in der Church of England vollzogen wurden und die Entwicklung der Mortalität von 1866 bis 1911 an. Zwischen diesen beiden Reihen beträgt der Korrelationskoeffizient 0.9512. Man könne nun annehmen, daß sich ein gemeinsamer Faktor für diese Gleichläufigkeit finden lasse (wie z. B. die Ausbreitung des wissenschaftlichen Denkens), doch die meisten Leser würden wohl mit ihm darüber übereinstimmen, daß es sich hierbei um eine “Nonsens-Korrelation” handele: “that it has no meaning whatever; that it is absurd to suppose that the two variables in question are in any sort of way, however indirect, causally related to one another.”³²²

³¹⁸ Yule (1921, S. 504). Vgl. auch Aldrich (1995, S. 371).

³¹⁹ “[...] correlation between movements may arise either because the one is causally dependent on the other, or because both are functions of the time, i. e., of some third variable or group of variables, on which both are causally dependent.” Yule (1921, S. 503).

³²⁰ Dies ist zu unterscheiden von der ebenfalls von ihm 1897 eingeführten ‘spurious correlation’, die auf den gemeinsamen Einfluß einer Drittvariable wie etwa dem technischen Fortschritt o. ä. zurückzuführen sei.

³²¹ Yule (1926, S. 4). Vgl. auch Aldrich (1995, S. 372). Der letzte Punkt ist, wie wir heute wissen, generell unzutreffend. Siehe etwa Granger/Newbold (1974).

³²² Yule (1926, S. 326).

Der Standardfehler des Korrelationskoeffizienten betrage bekanntlich $(1-r^2)/\sqrt{n}$, woraus sich sowohl ein im Vergleich zum Wert des Koeffizienten sehr geringer Standardfehler errechne als auch unter Annahme keiner Korrelation im "entire aggregate" ein sehr großer Wert für die "odds" ergeben würde, so daß nach herkömmlichen Maßstäben die Korrelation zweifelsfrei eine Bedeutung haben müsse. Der entscheidende Punkt sei jedoch, daß man in derartigen Situationen nicht von den Voraussetzungen ausgehen könne, die für die Berechnung des Standardfehlers nötig seien, insbesondere seien die einzelnen Werte keine unabhängigen Zufallsvariablen, sondern zeitlich abhängig voneinander.

Yule ging zunächst davon aus, daß den Daten eine harmonische Funktion zugrundeliegt. Abhängig von dem jeweiligen Ausschnitt, die in der konkreten Untersuchung zur Verfügung stehen, könne der Korrelationskoeffizient zwischen zwei Ausschnitten jeden Wert zwischen -1 und 1 annehmen.³²³ Diese Analyse dehnte er im folgenden auf "random series" aus: sowohl auf reine Zufallsreihen als auch auf aufsummierte Zufallsreihen sowie zweifach aufsummierte Zufallsreihen. Eine Unterscheidung der Reihen (in der Regel wisse man ja nicht, ob vorliegende Reihen "integriert" seien) könne anhand der seriellen Korrelationen durch die (Auto-) Korrelationsfunktion, deren Muster Aufschlüsse über den Typ der Reihe geben könne, ermittelt werden.³²⁴

Die Arbeit zeichnet sich durch eine Fülle von bemerkenswerten analytischen Einsichten aus, viele Anstöße erhielt Yule jedoch durch einfaches "Ausprobieren". Wenn das Problem sich analytisch als zu kompliziert oder aufwendig erwies, führte er einfach eine Reihe von Simulationen durch.³²⁵

Yules Arbeit von 1926 kann für die Entwicklung der Analyse konkreter Zeitreihen und der damit verbundenen Entwicklung von Konzepten als vorläufiger Abschluß angesehen werden. Bereits im folgenden Jahr publizierte Yule eine weitere klassische Zeitreihen-Untersuchung, die nun aber nicht mehr die Korrelation *zwischen* Zeitreihen, sondern *innerhalb* von Zeitreihen als Ausgangspunkt wählte. Diese Idee klang in seiner Arbeit von 1926 bereits an. Sie stellt eine Synthese aus zwei Entwicklungslinien dar; zum einen aus der besprochenen Analyse konkreter Zeitreihen, zum anderen stand sie aber auch (implizit³²⁶) in einer Tradition, die ihre Ursprünge bis in das 19. Jahrhundert zurückverfolgen konnte: die Theorie der stochastischen Prozesse.

³²³ Siehe ebda., Section II.

³²⁴ Interessanterweise finden inhaltliche Überlegungen (über "causes" etc.) hier keine Erwähnung mehr. Warum sollte z. B. der von ihm betrachtete *Beveridge-Wheat-Price-Index* eine aufsummierte "random series" sein?

³²⁵ Die genannten Simulationen erbrachten eine Normalverteilung für den ersten, eine annähernde Normalverteilung (mit abfallenden Enden) für den zweiten und eine U-förmige Verteilung für den dritten Fall. Bei Hendry/Morgan (1995b, S. 13ff) findet sich eine Wiederholung dieser Experimente anhand einer computergestützten Simulation. Sie weisen darauf hin, daß erst Phillips (1986) analytische Lösungen präsentierte.

³²⁶ Sie rezipiert diese Entwicklung nicht explizit, wird aber von der weiteren Entwicklung der Theorie der stochastischen Prozesse aufgenommen.

2. Die moderne Theorie der *stochastischen Prozesse*³²⁷ hat ihren Ursprung in der "Brownschen Bewegung".³²⁸ Der britische Botaniker Robert Brown [1773-1858] entdeckte 1827, daß die Bewegungen kleinster Teilchen in Flüssigkeiten oder Gasen keine biologischen, sondern physikalische Ursachen hatten.³²⁹ Überraschenderweise hat diese Beobachtung, abgesehen von einigen anfänglichen Erklärungsversuchen, keine größere Beachtung gefunden.³³⁰ Erst Jahrzehnte später widmete L. Gouy dem Problem im Zusammenhang mit der wieder zunehmenden Bedeutung der kinetischen Gastheorie stärkere Beachtung. Diese Arbeit wurde von H. Poincaré aufgegriffen; dessen Schüler L. Bachelier wiederum versuchte, anhand dieser Brownschen Bewegung eine Theorie der Spekulation zu entwickeln, die auf stochastischen Prozessen beruhen sollte.³³¹ Eine Erklärung wurde schließlich gegen Ende des 19. Jahrhunderts wiederum im Zusammenhang mit der kinetischen Gastheorie gefunden.³³²

³²⁷ Neben der reinen (mathematischen) Wahrscheinlichkeitstheorie wurden zur Entwicklung der Theorie stochastischer Prozesse Beiträge aus verschiedenen Gebieten der Physik (theoretische Physik, Geophysik, Quantenmechanik etc.), der Mechanik, der Astronomie, Meteorologie, Ozeanographie, Kontrolltheorie, Akustik, Elektrotechnik, Chemie, Biologie, Genetik, Medizin, Ökonomie, Soziologie, Demographie, um nur einige zu nennen, beigesteuert. Wir können hier nur einige ausgewählte Problemkreise darstellen. Eine systematische Darstellung der historischen Entwicklung der Theorie stochastischer Prozesse liegt noch nicht vor. Eine Bibliographie der Arbeiten von etwa 1900 bis 1959 liefert Wold (1965a). Diese Bibliographie, die etwa 6000 mit zusätzlichen Kriterien gekennzeichnete Titel umfaßt und von über 50 führenden Wissenschaftlern aus 30 Ländern zusammengestellt wurde (u. a. L. Schmetterer, A. Rényi, C. R. Rao, B. de Finetti, P. Whittle, H. Wold, K.-G. Jöreskog, H. Cramér, A. Kolmogorov, M. Bartlett, M. G. Kendall, J. Durbin, E. Parzen, K. Arrow), dürfte die maßgebliche Literatur erfaßt haben. Porter (1986, S. 193ff) behandelt die Entwicklung einiger statistischer Konzepte in der Physik, in denen das Problem der Zeit eine Rolle spielt (kinetische Gastheorie, statistische Mechanik; siehe dazu auch Schneider (1988, S. 301ff)), geht aber nicht explizit auf die Theorie der stochastischen Prozesse ein.

³²⁸ Im Gegensatz zur Komponentenvorstellung, die die Existenz von Komponenten a priori postuliert (die statistische Evidenz liefert hier nur Hinweise über die Art und Bedeutung der einzelnen Komponenten), steht bei der Theorie stochastischer Prozesse das verschiedene Zeitpunkte verbindende Gesetz im Mittelpunkt des Interesses.

³²⁹ "A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies." Zuerst veröffentlicht im *Edinburgh New Philosophical Journal* 5 (1828), S. 358-371. Zitiert nach Brush (1968, S. 349). Die Bewegungen wurden auch in Pflanzen festgestellt, die seit über hundert Jahren tot waren.

³³⁰ Vgl. zu dieser Entwicklung ausführlich Brush (1968).

³³¹ Diese frühe Adaption wahrscheinlichkeitstheoretischer Modelle auf ökonomische Phänomene hat freilich für die Entwicklung der Ökonometrie keine Rolle gespielt (s. u., Abschnitt B 2 b). Jevons hatte auch zwei Beiträge zur Theorie der Brownschen Bewegung beigesteuert, allerdings ohne Bezug auf seine ökonomischen Untersuchungen. Vgl. Brush (1968).

³³² Die mikroskopisch beobachtbaren Partikel setzen sich aus Atomgruppen (Molekülen) zusammen, die sich in permanenter Bewegung befinden. Die Verteilung der Geschwindigkeit ist dabei durch die Temperatur vorgegeben. Jedes fixierte Partikel kollidiert mit den es umgebenden Molekülen, und jede Kollision verändert die Geschwindigkeit des Partikels um einen kleinen Betrag. Der Nettoeffekt aller (kleinen) Veränderungen ist dann eine erratische, zufällige Bewegung des Partikels. Eine quantitative Theorie für dieses Phänomen wurde schließlich unabhängig voneinander von Albert Einstein und Marian Smoluchowski 1905 / 1906 entwickelt und im Hinblick auf die Modellierung stochastischer Prozesse in der Physik vor allem von N. Wiener und A. Kolmogoroff ausgebaut.

Ein zweiter Ursprung der Theorie stochastischer Prozesse läßt sich in der reinen Wahrscheinlichkeitstheorie finden, und zwar ohne daß hier ein konkretes Anwendungsproblem eine Lösung motivierte. Eine der großen Herausforderungen auf dem Gebiet der Wahrscheinlichkeitstheorie stellte um die Jahrhundertwende eine möglichst allgemeine Formulierung des Zentralen Grenzwertsatzes dar. A. Ljapunow hatte 1900 und 1901 einen Beweis präsentiert, der auf sehr schwachen Voraussetzungen beruhte.³³³ Dieser Beweis setzte unter anderem Unabhängigkeit der einzelnen Glieder voraus. A. A. Markow entwickelte, "herausgefordert durch den Erfolg von Ljapunow"³³⁴, auf der Suche nach noch allgemeineren Bedingungen daraufhin einen Beweis, der auch für voneinander abhängige Zufallsvariablen galt: diese nannte er "verkettete Größen".³³⁵ Markow war sich allerdings der fundamentalen Bedeutung, die seine Ausführungen für die Theorie stochastischer Prozesse und damit auch für zahlreiche Anwendungen haben sollten, nicht bewußt.³³⁶ "Erst langsam wurde man in der Folgezeit gewahr, daß eine Reihe von Problemen, etwa der Vererbungslehre oder der Diffusion in der Physik der Gase, mit dem von Markow angebotenen Begriff der Kette und den dafür entwickelten Methoden lösbar schien."³³⁷ Markow selbst hat im folgenden seinen Beweis auf Abhängigkeiten höherer Ordnung ausgedehnt,³³⁸ jedoch nur ein konkretes Beispiel für die Anwendungsmöglichkeit dieser abhängigen Zufallsvariablen angegeben.³³⁹ In der Folgezeit haben jedoch nicht diese, sondern G. U. Yules Arbeit über AR(2)-Prozesse sowie E. Slutzkys Untersuchung über die Auswirkungen der Summation of Random Causes den Grundstein für die moderne Theorie stochastischer Prozesse bestimmt.³⁴⁰

³³³ Vgl. Schneider (1988, S. 120).

³³⁴ Ebda., S. 443. Beide waren Schüler des russischen Mathematikers P. L. Tschebyschew und standen in einem Konkurrenzverhältnis zueinander. Wissenschaftssoziologische Faktoren sind bei der Entwicklung der Theorie stochastischer Prozesse auch im folgenden von Bedeutung. Aufschlußreich ist hier die Autobiographie von Norbert Wiener (1962).

³³⁵ Andrej Andrejewitsch Markow, *Die Erweiterung des Gesetzes der großen Zahlen auf Größen, die voneinander abhängig sind* (russisch), in: *Abhandlungen der physikalisch-mathematischen Gesellschaft der Universität Kasan* 15/2 (1906), S. 135-156, zitiert nach Schneider (1988, S. 443). Später wurden diese nach ihm "Markow-Ketten" genannt.

³³⁶ Das betonte Neyman (1981, S. vii).

³³⁷ Schneider (1988, S. 444).

³³⁸ A. A. Markow, *Über verbundene Größen, die keine eigentlichen Ketten bilden* (russisch), in: *Mitteilungen der Petersburger Akademie der Wissenschaften* 5/6 (1911), S. 113-126, in deutscher Übersetzung in A. A. Markoff, *Wahrscheinlichkeitsrechnung* (übersetzt von H. Liebmann), Leipzig/Berlin (1912), S. 299-311, nach Schneider (1988, S. 444).

³³⁹ Bei diesem Beispiel handelte es sich um die Frage, wie man die Wahrscheinlichkeiten bestimmt, daß auf Konsonanten Vokale folgen und vice versa. Markow untersuchte dies anhand des Romans "Eugen Onegin" von M. Puschkin: A. A. Markow, *Versuche einer statistischen Untersuchung über den Text des Romans 'Eugen Onegin' zur Beleuchtung des Zusammenhangs der Kettenversuche* (russisch), in: *Mitteilungen der Petersburger Akademie der Wissenschaften* 7/6 (1913), S. 153-162, zitiert nach Schneider (1988, S. 444).

³⁴⁰ In den zwanziger Jahren erscheinen eine Reihe von zeitreihenanalytischen Beiträgen im Zentralorgan der amerikanischen statistischen Gesellschaft, dem *Journal of the American Statistical Association*, vor allem über Saisonbereinigung, Glättung und Trendbereinigung ökonomischer Reihen. Diese Arbeiten haben jedoch für die Entwicklung der statistischen Zeitreihenanalyse keine weitere Rolle gespielt. Siehe dazu unten, Abschnitt B 2 b.

Unmittelbar nach seiner Studie über "Nonsense-Korrelationen" stellte Yule die interne Struktur in einem univariaten Zusammenhang in den Vordergrund seiner Untersuchungen und verknüpfte dies mit den nach seiner früheren Ansicht wichtigen oszillatorischen Komponenten. Nun stellte er jedoch die "Restkomponente" in den Mittelpunkt des Interesses. Bis zu diesem Zeitpunkt gingen Untersuchungen von univariaten Zeitreihen davon aus, daß für diese ein "Komponentenmodell" angemessen sei (s. o.), bei dem zufällige Störungen Komponenten mit spezifischem Bedeutungsgehalt überlagerten. Yule ging dagegen von einer völlig anderen Vorstellung aus. Er schrieb dazu zunächst eine harmonische (Sinus-) Funktion in Form einer *Differenzengleichung*.³⁴¹

$$(8) \quad y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t.$$

In dieser Darstellung, die ohne Zufallseinflüsse mit der funktionalen Darstellung völlig identisch war, ergab sich jedoch eine fundamental andere Rolle für die Zufallsgrößen. Yule beschrieb das Verhalten einer derart gestörten Reihe mit einer mittlerweile berühmten Analogie: Die Bewegung eines Pendels werde in gleichabständigen Intervallen gemessen, wodurch eine reine trigonometrische Schwingung beschrieben wird. Diese *Messungen* seien durch nicht perfekte Meßinstrumente fehlerbehaftet, rein additiv und unabhängig voneinander. Dann passiere folgendes:

"The recording apparatus is left to itself, and unfortunately boys get into the room and start pelting the pendulum with peas, sometimes from one side and sometimes from the other. The motion is now affected, not by *superposed fluctuations* but by true *disturbances*, and the effect on the graph will be of an entirely different kind. The graph will remain surprisingly smooth, but amplitude and phase will vary continually."³⁴²

Wenn man eine Sinus-Funktion in Form einer Differenzengleichung schreiben konnte, so war andererseits die Differenzengleichungsform nicht auf eine bestimmte Form beschränkt. Yule untersuchte daher auch Modelle mit drei anstelle von zwei zeitversetzten "Regressoren" und betonte, daß die Anzahl dieser zeitversetzten Größen prinzipiell nicht beschränkt sei. So könne man ganz allgemein Modelle der Form

$$(9) \quad y_t = \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t$$

betrachten und deren Parameter mittels des Kleinstquadratverfahrens bestimmen.³⁴³ Mit dieser Modellierung waren die sogenannten "autoregressiven Prozesse" (AR(*p*)-Modelle) neugeboren, die Markow 1906/1911 aus rein mathematischen Erwägungen abgeleitet hatte.

Als eine weitere Grundlage der Theorie stochastischer Prozesse gilt die 1927/1937 erschienene Arbeit von E. Slutsky über die *Summation of Random Causes*.³⁴⁴ Im

³⁴¹ Die technischen Details finden sich in Rahlf (1996b).

³⁴² Yule (1927 [1971, S. 390]). Hervorhebungen im Original.

³⁴³ Rechentechnisch wurde die Bestimmung der Parameter von Walker (1931) verfeinert.

³⁴⁴ Eugen Slutsky [1880-1948] studierte in Kiew Physik und Mathematik, in München Ingenieurwissenschaften und graduierte 1911 an der rechtswissenschaftlichen Fakultät in Kiew.

Gegensatz zu Yule ging Slutsky nicht von zeitverzögerten beobachteten Werten y_{t-i} als "Ursache" aus, sondern von einer Kombination von externen, nicht direkt beobachtbaren Zufallsvariablen. Er untersuchte Modelle der Form³⁴⁵

$$(10) \quad y_t = \sum_{i=1}^q \theta_i \varepsilon_{t-i},$$

die später als "moving-average"- bzw. MA(q)-Modelle bezeichnet wurden. Slutsky stellte zu Beginn seiner Arbeit fest, daß nahezu alle ökonomischen wie auch viele andere uns umgebende Phänomene sich in wellenartigen Auf- und Abbewegungen vollziehen. Wie die Wellen des Meeres wiederholen sich diese jedoch nicht auf identische Weise, sondern variieren sowohl in der Amplitude als auch in der Periode. Dennoch werde man in der Lage sein, allgemeine Regelmäßigkeiten festzustellen, Summen von Wellen verschiedener Größe zu sehen,³⁴⁶ die eine harmonische Analyse nahelegen.³⁴⁷ Falls mit dieser Methode keine völlige Übereinstimmung erzielt werden könne, so seien die Abweichungen als unbedeutende Störeinflüsse, "superimposed on the regular waves" (S. 105), aufzufassen.³⁴⁸ Auch divergierende Ergebnisse für Teilbereiche würden diese Vorgehensweise nicht grundsätzlich in Frage stellen. Dennoch sei man in dieser Situation mit zwei Problemen konfrontiert: So setze die harmonische Analyse voraus, daß die einzelnen Elemente voneinander unabhängig seien; empirische Reihen seien aber in der Regel durch serielle Abhängigkeit gekennzeichnet:

"Thus we are led to our basic problem: is it possible that a definite structure of a connection between random fluctuations could form them into a system of more or less regular waves?"³⁴⁹

Eine weitere Schwierigkeit ergäbe sich, wenn man schließlich solche Regelmäßigkeiten entdeckt habe:

1913 wurde er Mitglied der dortigen Fakultät am Handels-Institut, 1929 ebda. Professor. 1926 wechselte er zum Moskauer Konjunkturinstitut, 1934 zum mathematischen Institut der Moskauer Universität, 1936 zum mathematischen Institut der Moskauer Akademie der Wissenschaften. Dort blieb er bis zu seinem Tod. Slutsky war Mathematiker, Statistiker und Ökonom; in der Ökonomie hat er einen bedeutenden Beitrag zur Konsumtheorie geleistet. Siehe dazu Gandolfo (1990, S. 259f). Auf dem Gebiet der mathematischen Statistik entwickelte er mehrere Jahre vor R. A. Fisher die später auf diesen zurückgeführte χ^2 -Minimum-Methode (1914). Die hier behandelte Arbeit gilt als eine der Grundlagen der Theorie stochastischer Prozesse, da unter anderem mehrere wichtige Theoreme bewiesen werden. Sie erschien erstmals 1927 in russischer Sprache mit englischer Zusammenfassung und fand auch im Westen schnell Verbreitung. Eine erweiterte Fassung erschien 1937 in englischer Übersetzung auf Initiative von H. Schultz und R. Frisch in der *Econometrica*. Vgl. Slutsky (1937, hier S. 105).

³⁴⁵ Für die genaue Form der Modelle siehe Rahlf (1996b).

³⁴⁶ "The eye of the observer instinctively discovers on waves of a certain order other smaller waves, so that the idea of harmonic analysis, viz., that of the possibility of expressing the irregularities of the form and the spacing of the waves by means of the summation of regular sinusoidal fluctuations, presents itself to the mind almost spontaneously." Slutsky (1937, S. 105).

³⁴⁷ Er bezieht sich hier auf die 1898 von Schuster eingeführte Periodogrammanalyse.

³⁴⁸ Es handelt sich hierbei also nicht um die von Yule vorgeschlagenen AR(2)-Modelle.

³⁴⁹ Slutsky (1937, S. 106).

“What is the mechanism of causality which, decade after decade, reproduces the same sinusoidal wave which rises and falls on the surface of the social ocean with the regularity of day and night. [...] The presence of waves of definite orders, the long waves embracing decades, shorter cycles from approximately five to ten years in length, and finally the very short waves, will always remain a fact begging for explanation. The approximate regularity of the periods is sometimes so distinctly apparent that it, also, cannot be passed by without notice. Thus in short, *the undulatory character of the processes and the approximate regularity of the waves* are the two facts for which we shall try to find a possible source in random causes combining themselves in their common effect.”³⁵⁰

Die methodologische Vorgehensweise zur Klärung dieser Fragen sei sowohl deduktiv als auch induktiv. Nur wenige Probleme hätten deduktiv gelöst werden können, eine Reihe von Erkenntnissen sei durch “Experimente” erfolgt: “It is experiment, that shows us totally unexpected facts, thus pointing out problems which otherwise would hardly fall within the field of the investigator.”³⁵¹

Die Arbeiten von Yule und Slutsky waren für die weitere Entwicklung der Theorie der stochastischen Prozesse konstituierend. In der Folgezeit wurden vor allem technische Spezialprobleme – in diesem vorgegebenen Rahmen – gelöst. Die inferenzstatistische Behandlung des Zyklusproblems spielte dabei keine große Rolle mehr. Wold bemerkte hierzu noch 1938:

“However, it should be observed that the relevant problem does not consist merely in calculating the variance or the distribution of a single autocorrelation coefficient. We have also to face the much deeper question concerning the reliability of the periodicities which present themselves in the graph of the serial coefficients. In view of the complications already occurring in sampling problems involving merely one individual serial coefficient, the possibility of arriving at a practicable, quantitative measure of significance in this connexion seems, at least for the moment, hopeless.”³⁵²

Wold selbst verwendete zwar auch Beispiele aus dem Bereich der Ökonomie, doch wurde seine Arbeit nicht von der Ökonometrie, sondern von der theoretischen Statistik, insbesondere von Kolmogoroff im Rahmen der mathematischen Wahrscheinlichkeitstheorie, aufgegriffen.³⁵³

³⁵⁰ Ebda., S. 106f. Hervorhebung im Original.

³⁵¹ Ebda., S. 107. Er sah seine Arbeit sowohl in inhaltlicher als auch methodologischer Hinsicht in einer Linie mit Yule (1926). Es wäre reizvoll, die Arbeiten von Yule und Slutsky systematisch zu vergleichen, dies muß jedoch einer späteren Arbeit vorbehalten bleiben. Einige technische Aspekte hinsichtlich der Beziehung beider Modelltypen finden sich in Rahlf (1996b).

³⁵² Wold (1938, S. 109). Das Problem wurde, soweit wir sehen, erst wieder von Thome (1994a) als solches erkannt und von uns in Rahlf (1996b) in den Mittelpunkt des Interesses gestellt. Eine erste umfassende Formalisierung zeitreihenanalytischer Verfahren unter Berücksichtigung des Zeit- und Frequenzbereiches veröffentlichte Norbert Wiener 1930, in der die harmonische Analyse mit der Theorie stochastischer Prozesse verbunden wurde, siehe Wiener (1930). Brillinger (1984, S. li) bezeichnet Wiener als “the individual whose work had dramatic effect on time series practice” in den vierziger Jahren. Yule (1915) hatte in einer Besprechung von Moore (1914) noch dessen Verwendung des Periodogramms für ökonomische Daten kritisiert, da diese keine feste Periode beinhalteten. In Yule (1927) fanden sich hinsichtlich der Periodogrammanalyse der von ihm vorgestellten AR(2)-Prozesse lediglich “some trials”, bei Slutsky (1937) spielte es keine Rolle.

Seit den fünfziger Jahren sind die Beiträge zur Theorie stochastischer Prozesse nur noch rein mathematischer Natur.³⁵⁴ Für den Anwendungsbereich wurde 1970 die Monographie von G. E. P. Box und G. M. Jenkins zu einem "Meilenstein".³⁵⁵ Box und Jenkins griffen die von Yule und Slutsky eingeführten und mittlerweile weiterentwickelten Modelle auf und verbanden sie zu einer Synthese. Zwar konnte unter bestimmten Bedingungen ein autoregressiver Prozeß in einer "moving-average"-Form dargestellt werden, andererseits ein moving-average-Prozeß in autoregressiver Form,³⁵⁶ doch hatten diese alternativen Formen eine Vielzahl von Parametern, die eine Bestimmung anhand der Daten erschwerten. Box und Jenkins faßten (9) und (10) zu einer Modellklasse von autoregressiven moving-average-Modellen (ARMA-Modellen) zusammen:

(11)

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

Ziel war es dann, für gegebene Daten ein möglichst sparsames Modell zu finden, somit also die Ordnung des Modells (p und q) zu bestimmen und die Parameter des bestimmten Modells zu schätzen. Auch trendbehaftete Reihen konnten mit diesen Modellen behandelt werden; in solchen Fällen sollten anstelle der Originalreihe deren Differenzen betrachtet werden.³⁵⁷ Weiterhin wurde dieses Grundgerüst

³⁵³ "My theory did not influence econometrics at the time; this is my understanding now. People were interested, but it was not followed up by other econometricians, so I was known at the time for my work on demand." Wold in Hendry/Morgan (1994, S. 422). Herman Ole Andreas Wold [1912-1992] war gebürtiger Norweger, seine Familie ging jedoch bereits 1912 nach Schweden. Seit 1927 studierte er an der Universität von Stockholm unter anderem bei Harald Cramér, arbeitete anschließend für eine Versicherungsgesellschaft und promovierte 1938 mit der erwähnten Arbeit *A Study in the Analysis of Stationary Time Series*. Bis 1942 blieb er als Dozent für mathematische Statistik in Stockholm und erhielt dann den Lehrstuhl für Statistik an der Universität Uppsala. Dort richteten sich seine Interessen auf die Ökonometrie, insbesondere auf die Nachfrageanalyse und Kausalkettenmodelle, später auf das von ihm so genannte *soft modelling*, ein Ansatz, der im folgenden von einem seiner Schüler, K. G. Jöreskog, zu dem in der Soziologie mittlerweile sehr verbreiteten LISREL-Ansatz ausgebaut wurde. 1970 wechselte Wold auf den Lehrstuhl für Statistik der Universität Göteborg und blieb dort bis zu seiner Emeritierung 1975. Eine Bibliographie seiner Arbeiten findet sich in Hendry/Morgan (1994).

³⁵⁴ Ein gutes Beispiel ist die Monographie von Doob (1953). Für weitere Belege siehe Wold (1965a). Auf die – für die theoretische Seite bedeutende – Entwicklung der Spektralanalyse sowie die – für die praktische Seite bedeutende – Entwicklung der exponentiellen Glättung wollen wir hier nicht eingehen. Die multivariate Analyse stand zu dieser Zeit eher im Hintergrund. Siehe dazu die Arbeit von Quenouille (1957).

³⁵⁵ Box/Jenkins (1976). Die erste Auflage des Buches erschien 1970. George E. P. Box [1919-] ist ein Schwiegersohn von R. A. Fisher. Er studierte Chemie an der London University. Von 1948 bis 1956 arbeitete er als Angestellter der *Imperial Chemical Industries*, von 1957 bis 1959 war er (auf Initiative John Tukeys) Direktor der *Statistical Techniques Research Group* an der *Princeton University*, seit 1960 Professor an der Universität Wisconsin ("R. A. Fisher professor of statistics"), wo er das *Department of Statistics* gründete. Ein Teil seiner Schriften ist in Tiao (1984) wiederabgedruckt. Gwilym M. Jenkins [1934-1982] promovierte nach einem Studium der Ingenieurwissenschaften 1956 am *University College*, war 1959/60 *visiting professor* in Stanford und Princeton, seit 1965 Professor für *systems engineering* an der Universität Lancaster. 1974 verließ er die Universität, um eine ingenieurwissenschaftliche Beratungsfirma zu gründen.

³⁵⁶ Siehe dazu Thome (1994a, S. 24, 32).

auf Modelle für saisonale Reihen sowie um eine explizite Berücksichtigung von exogenen Variablen in sogenannten Transferfunktionen erweitert.

Gegenüber den Arbeiten von Yule, Slutsky, Wold und anderen waren die von Box und Jenkins vorgestellten univariaten Modelle keine wesentliche Erweiterung. Daß dennoch von einer “Box-Jenkins-Methodologie” gesprochen werden kann, hängt vor allem mit der ebenfalls von ihnen vorgeschlagenen Modellierungsstrategie zusammen, der “general philosophy of diagnostic checking”.³⁵⁸

Das Schema dieser Vorgehensweise nach Box/Jenkins ist mittlerweile in zahlreichen Lehrbüchern und Einführungstexten – mehr oder weniger modifiziert – übernommen worden:

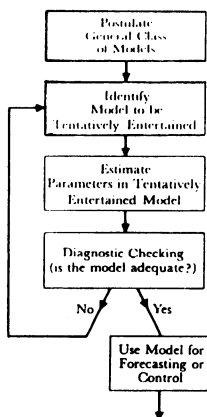


Abb. 3: Stadien des iterativen Ansatzes der Modellbildung nach Box/Jenkins. Quelle: Box/Jenkins (1976, S. 19).

Dieses iterativ zu durchlaufende, originale Schema beginnt mit einem Schritt, der in vielen Adaptionen in Lehrbüchern zumeist unterschlagen wird: *Postuliere* eine generelle Klasse von Modellen. Box und Jenkins gingen zumindest nach diesem Schema also keineswegs davon aus, daß die von ihnen vorgeschlagenen ARIMA-Modelle universal anzuwenden seien.

In einem weiteren Schritt ist ein spezielles Modell aus der Klasse aller möglichen Modelle auszuwählen. Dies geschieht über die Autokorrelationen und partiellen Autokorrelationen einer Reihe. Die Autokorrelationsfunktionen und partiellen Autokorrelationsfunktionen von ARMA-Prozessen weisen charakteristische Muster auf. Ein Vergleich dieser Muster mit den für eine konkrete Reihe errechneten Werten soll hier eine “Identifikation”³⁵⁹ ermöglichen, soll es erlauben, aus der Klasse der möglichen Modelle das “wahre” Modell herauszufinden.

³⁵⁷ In theoretischer Hinsicht bedeutete eine solche Vorgehensweise, daß der zugrundeliegende Prozeß “integriert”, d. h. aufsummiert sein mußte. Box und Jenkins sprachen dann von “autoregressive integrated moving average (ARIMA-) models”.

³⁵⁸ Box (1980, S. 390). Diesen Punkt übersieht etwa Priestley in seinem Kommentar zu Chatfield/Prothero (1973, S. 331), in dem er den Ausdruck “Box-Jenkins method” als irreführend bezeichnet, da die Techniken bereits von anderen eingeführt worden seien.

Ob dies, von technischen Schwierigkeiten einmal abgesehen, *überhaupt* möglich ist, wurde von Kritikern bezweifelt. So hat etwa M. Kendall in einer Besprechung des Buches kritisiert, daß in vielen Situationen zwischen einzelnen Modellen nicht hinreichend diskriminiert werden könne. Box und Jenkins antworteten auf diese Kritik mit einem Beispiel, das die folgenden MA(1)- und AR(1)-Modelle gegenüberstellte,³⁶⁰

$$(12) \quad \begin{aligned} y_t &= \varepsilon_t - 0.2\varepsilon_{t-1} \\ y_t &= -0.2y_{t-1} + \varepsilon_t, \end{aligned}$$

die ein sehr ähnliches Verhalten aufweisen, aber inhaltlich völlig unterschiedlich zu interpretieren seien. Sie fragten: "How can the two models be distinguished? For the series of lengths we are likely to examine, the answer is, of course, that they cannot. *But then they do not need to be.* Either model would fit the data equally well and it is clear that it does not matter which we use. *These models, like every other, are approximations.* When two approximations are essentially equivalent we can use either and obtain essentially the same result."³⁶¹

Für uns ist in diesem Zusammenhang die Inferenzkonzeption der Autoren von besonderer Bedeutung. Vergegenwärtigt man sich die "Vorgeschichte" beider Autoren,³⁶² so ist die in ihrem Buch nicht weiter erläuterte Einbettung in einen häufigkeitstheoretischen Ansatz auf den ersten Blick geradezu verblüffend. Es findet sich zwar ein kurzer Abschnitt "Estimation using Bayes' theorem" (S. 250-258), dieser steht jedoch in keinerlei Zusammenhang mit dem übrigen Text und ist wie einige andere Abschnitte durch einen senkrechten Strich gekennzeichnet, "to indicate material which may be omitted in the first reading" (S. xi). Das Likelihood-Prinzip fand nun nur noch kurze Erwähnung. Zwar wurde ausführlich auf die Ableitung der Likelihoodfunktionen für die Parameter der verschiedenen Modelle eingegangen, auch ihre graphische Darstellung wurde wiederholt angeboten und empfohlen, wie auch eine graphische Darstellung der Residuenquadratsummen. Die vorhergehende Identifikation des Modells (S. 173ff) findet jedoch ebenso wie die sich an die Schätzung anschließende Modelldiagnose (S. 285ff) anhand von Signifikanztests statt.³⁶³

³⁵⁹ Schlittgen/Streitberg (1994, S. 288) sprechen in diesem Zusammenhang zu Recht von "Spezifikation" anstelle von "Identifikation".

³⁶⁰ Vgl. Box/Jenkins (1973, S. 341).

³⁶¹ Ebda. Hervorhebungen von uns.

³⁶² Box hatte 1973 zusammen mit G. C. Tiao eines der ersten Lehrbücher zur Bayes-Statistik publiziert. Auf diese Konzeption kommen wir im nächsten Abschnitt noch einmal zu sprechen. G. W. Jenkins hatte 1962 zusammen mit C. B. Winsten und G. A. Barnard einen grundlegenden Beitrag zur Übertragung des Likelihood-Prinzips auf Zeitreihen veröffentlicht und noch 1969 zusammen mit D. G. Watts in der zweiten Auflage ihres Buches *Spectral analysis and its applications* in einem umfangreichen ersten Kapitel das Likelihood-Prinzip als im Vergleich zur *sampling-theory* angemessenere Inferenz beschrieben. Siehe Box/Tiao (1973), Barnard/Jenkins/Winsten (1962), Jenkins/Watts (1969).

³⁶³ Entsprechend ist auch Teil III (Transferfunktionsmodelle) aufgebaut, der allerdings lediglich Kleinstquadratschätzungen und keine Likelihoodfunktionen behandelt.

Aufschlußreich ist in diesem Zusammenhang, wie das Werk entstand. Über die Genese des Buches finden wir einige Bemerkungen von Box in einem Interview aus dem Jahre 1987. Box begann, sich für die Zeitreihenanalyse im Zusammenhang mit einem Temperaturregelungsproblem zu interessieren.³⁶⁴ Bei der Konstruktion eines Katalysators, in dem Temperaturregelungen mittels Sinus-Funktionen gesteuert wurden, machte Jenkins Box auf eine Reihe von Problemen bei der Messung der Wirkung dieses Inputs aufmerksam:

“He said ‘When you change the temperature by putting a sine wave into it there will be a delay in the system because of the mixing and the reaction, and so you’ll get a delayed and attenuated sine wave coming out. You’ll get a shift in the phase, and all this kind of stuff. And not only that, but if you don’t take proper account of the serial correlation in the noise, you’ll get more wrong answers.’ So that’s what we started to work on.”³⁶⁵

Die Lösung der in dieser konkreten Situation entstandenen Probleme schuf weitere:

“We had no intention on working on time series. We were working on optimization to begin with, and then gradually we realized that it was a control problem. Finally we realized that control involved forecasting because you can regard simple control algorithms as forecasting how big the deviation will be at the end of the next interval and then taking an action which cancels out the forecast deviation. We started looking at methods that had been used in forecasting; the ones that seemed to work well were things like exponential smoothing, which implied the importance of particular kinds of nonstationary time series models, which we then worked on.”³⁶⁶

Deutlicher läßt sich wohl die pragmatische Behandlung eines Problems nicht beschreiben. Betrachtet man hingegen das Buch, ist diese Genese nicht mehr zu erkennen. Das Kontrollproblem findet sich am Schluß nach einer über 400seitigen Behandlung der Eigenschaften und Analysemöglichkeiten stochastischer Prozesse. Auch hierfür lieferte Box 1987 eine Erklärung:

“The book that finally came out [...] is sort of backward compared to the way we got in. The control part is at the end and I don’t think that the actual problem we started with, pursuing the maximum, even gets mentioned in the book. But that’s the place we actually started. We worked back and then we realized we had to do something about nonstationary time series in order to do that. So that was the way that book evolved.”³⁶⁷

Wie erklärt sich dann aber der große Erfolg des Buches auch und gerade in “nichttechnischen” Fachwissenschaften wie der Ökonomie? Die Anwendungsmöglichkeiten der dort beschriebenen Verfahren wurden nicht auf Probleme wie Temperaturregelungen in Katalysatoren beschränkt, sondern es wurden bereits in der Einleitung folgende “Probleme” thematisiert:

³⁶⁴ “My interest in it was very oblique and pertained to optimum conditions. I had met some problems in my consulting where the maximum was actually moving.” Box in DeGroot (1987, S. 251).

³⁶⁵ Ebda.

³⁶⁶ Ebda.

³⁶⁷ Ebda.

1. Prognose in
 - (a) economic and business planning³⁶⁸
 - (b) production planning
 - (c) inventory and production control
 - (d) control and optimization of industrial processes
2. Schätzung von Transferfunktionen im Zusammenhang mit industriellen Prozessen
 - (a) to achieve better control of existing plants, and
 - (b) to improve the design of new plants
 und allgemein zur Prognose (ohne konkrete Beispiele).
3. Design diskreter Kontrollsysteme, die Ingenieuren eine Steuerung von Prozessen ermöglichen sollten.

Die postulierten Anwendungsmöglichkeiten umfaßten also einen viel weiteren Bereich als denjenigen, aus dem ursprünglich die Idee des Buches geboren wurde, wobei die Bedeutung für ökonomische Anwendungen lediglich im Bereich des "Forecasting" gesehen wurde.

Die "Box-Jenkins-Methodologie" fand seit Ende der siebziger Jahre viele Anhänger.³⁶⁹ Dabei stand in dieser frühen Phase noch stärker als heute der Prognose-Aspekt im Vordergrund, da Prognosen univariater ARIMA-Modelle in vielen Fällen den herkömmlichen Verfahren (Trendextrapolationen, exponentiellem Glätten) überlegen zu sein schienen.³⁷⁰

In den bislang betrachteten Ansätzen wurde stets davon ausgegangen, daß zwischen einzelnen Zeitpunkten zwar Abhängigkeiten bestehen konnten, die Parameterstruktur selbst wurde dabei aber im Verlauf der Zeit als konstant angesehen. Wir wollen zum Abschluß dieses Kapitels noch auf eine Entwicklung hinweisen, die genau diese Annahme in Frage stellte.

Falls die Parameter eines Modells nicht konstant sind, können drei Situationen unterschieden werden:³⁷¹

- Die Struktur vor und nach einem bestimmten *Ereignis* kann sich unterscheiden,
- die Statistiken/Parameter können sich über die Zeit *deterministisch* oder (realistischer) *stochastisch entwickeln*,

³⁶⁸ Das erste konkrete Beispiel, das sie erwähnten, ist ein Problem der Kontrolle von Inventuren, ein "sales forecasting problem". Box/Jenkins (1976, S. 1).

³⁶⁹ Anderson (1977b, S. 285) spricht in diesem Zusammenhang sogar von der "Box-Jenkins-religion". In der anfänglichen Rezeption finden sich allerdings auch kritische Stimmen. Siehe z. B. die Diskussion in Chatfield/Prothero (1973).

³⁷⁰ Einige Artikel behandelten sogar explizit das Problem der Motivation von ARIMA-Modellen für "Praktiker" (= Manager), die aufgrund der scheinbar hohen formalen Anforderungen zögerten, diese Modelle in der Praxis anzuwenden. Vgl. zum Beispiel Anderson (1977a), Anderson (1977b).

³⁷¹ Vgl. Ledolter (1989).

- andererseits kann die Struktur zwar konstant, aber durch das gewählte Modell nicht adäquat erfaßt sein; dadurch kann eine *scheinbare* Parameterinkonstanz entstehen.

In der Statistik sind eine Reihe von Verfahren entwickelt worden, um diese Probleme anzugehen. Wenn von einer Strukturveränderung aufgrund eines Ereignisses ausgegangen wird, wird dies in der klassischen Statistik z. B. durch einen Test auf eine "signifikante" Veränderung evaluiert. Im Bereich der univariaten ARIMA-Modelle hatten Box und Tiao dafür die Anwendung von Interventionsmodellen empfohlen. Ist der Zeitpunkt dieses Wandels unbekannt, so kommt die Schätzung dieses Zeitpunktes zwar als zusätzliche Komplikation hinzu, an der grundsätzlichen Vorgehensweise ändert sich dadurch jedoch nichts. Andererseits können sich Parameter eines Modells kontinuierlich über die Zeit verändern. Falls man nicht von einer deterministischen Veränderung ausgeht, ist dann eine Verteilung und ein stochastisches Gesetz zu unterstellen, das die Entwicklung der Parameter über die Zeit erfaßt.

Einen zentralen Punkt hinsichtlich der Untersuchung von Regressionsmodellen auf Parameterkonstanz stellt die Arbeit von Brown/Durbin/Evans (1975) dar. Die behandelten Probleme wurden zwar schon vorher in einzelnen Aspekten thematisiert, Brown, Durbin und Evans präsentierten jedoch einen systematischen Ansatz und verbanden damit vor allem eine datenanalytische Konzeption.

Die Autoren wiesen gleich zu Beginn ihrer Arbeit darauf hin, daß ihr Ansatz zwar formale Signifikanztests enthalte, ihre "Philosophie" jedoch in Einklang mit der von Tukey (1962) entwickelten Datenanalyse stehe:

"Essentially, the techniques are designed to bring out departures from constancy in a graphic way instead of parametrizing particular types of departure in advance and then developing formal significance tests intended to have high power against these particular alternatives. From this point of view the significance tests suggested should be regarded as yardsticks for the interpretation of data rather than leading to hard and fast decisions."³⁷²

Sie betrachteten das Regressionsmodell

$$(13) \quad y_t = x_t' \beta_t + u_t, \quad t = 1, \dots, T$$

wobei die k Regressoren als nichtstochastisch angesehen wurden. Autoregressive Modelle sind damit ausgeschlossen. Die Fehlerterme wurden als unabhängig, normalverteilt mit Varianz σ_t^2 , $t = 1, \dots, T$ und nicht autokorreliert vorausgesetzt.³⁷³

Zur Überprüfung auf konstante Koeffizienten empfehle sich die Verwendung von transformierten rekursiven³⁷⁴ kumulativen Residuen w_t .³⁷⁵ Sie schlugen darauf

³⁷² Brown/Durbin/Evans (1975, S. 149f).

³⁷³ Vgl. ebda., S. 151.

³⁷⁴ Der Ausdruck "rekursiv" ist hier irreführend, da stets neue Beobachtungen in die Berechnung einfließen. Nelder schlug daher vor, das Verfahren als sequentiell oder als updating-Verfahren zu bezeichnen. Die Bezeichnung "rekursiv" hat sich dennoch durchgesetzt.

³⁷⁵ Eine graphische Darstellung der originalen (OLS-) Residuen einer konstanten Schätzung ist ihrer Ansicht nach bei Vorliegen variabler Koeffizienten nicht hilfreich, wenn sich die Koeffizienten β langsam oder graduell verändern.

aufbauend fünf verschiedene Verfahren vor.³⁷⁶ Der erste Test bestand aus einer graphischen Darstellung der kumulierten Größen w , gegen r , der zweite aus einer graphischen Darstellung der kumulierten Summen der quadrierten Größen. Als dritte Variante empfahlen sie eine gleitende Regression ("moving regression"); eine vierte Möglichkeit untersuchte den Fall, daß die Regressionskoeffizienten in der Zeit einem Trendpolynom folgen. Die letzte Variante schließlich untersuchte die Möglichkeit eines Strukturbruchs.

Die Berechnung dieser Größen sahen Brown, Durbin und Evans als ersten Schritt: *Wenn* die Darstellungen auf eine zeitliche Inkonzistenz hindeuten, sollte man schließlich zusätzlich die Pfade der einzelnen Elemente des geschätzten Koeffizientenvektors sowie der geschätzten Varianz graphisch untersuchen.³⁷⁷ Weitergehende Untersuchungen wurden vorerst nicht empfohlen.

In der sich anschließenden Diskussion, die dem Beitrag eine große Relevanz bescheinigte, wurden vor allem zwei Aspekte Gegenstand der Kritik: Zum einen wurde verständlicherweise von mehreren Diskussionsteilnehmern die eingeschränkte Anwendbarkeit der Verfahren durch die starken Annahmen von nicht-stochastischen Regressoren und nicht autokorrelierten Residuen kritisiert, da beide Phänomene in einer Reihe von realen Situationen anzutreffen seien.³⁷⁸ Für solche Fälle seien die angegebenen Testgrenzen in der Regel nicht anwendbar. Diese – technischen – Vorbehalte fallen zum anderen mehrfach zusammen mit einer grundsätzlichen Kritik des Ansatzes:

Die von Brown, Durbin und Evans vorgeschlagene Verwendung von Signifikanzgrenzen als Orientierungshilfe ohne "hard and fast decisions" wurde dabei sowohl aus der klassischen als auch aus einer bayesianischen Sichtweise kritisiert. So forderte etwa Phillips nähere Untersuchungen über die Macht der Tests gegenüber einer Reihe von spezifischen Alternativhypothesen, und Ehrenberg sah in der Datenanalysekonzeption überhaupt keinen Erkenntnisgewinn.³⁷⁹ Auch Harvey empfand den Verzicht auf spezifische Alternativhypothesen als unbefriedigend und forderte konkrete Alternativen, die machtvollere Tests erlaubten.³⁸⁰ Smith setzte dieser Unbestimmtheit eine bayesianische Argumentation entgegen:

"As they stand, the procedures put forward by the authors seem to require a great deal of informal use of personal judgement (as evidenced in Section 2.4, for example, where we encounter such phrases as 'we ourselves prefer', 'it may be useful to examine' and 'it is often informative to').

³⁷⁶ Die technischen Einzelheiten finden sich in etwa in Ledolter (1989).

³⁷⁷ Vgl. Brown/Durbin/Evans (1975, S. 155).

³⁷⁸ Vgl. ebda. die Diskussionsbeiträge von Kendall, Priestley, Young, Smith, Ehrenberg, Harvey, Khan, Quandt und Tong.

³⁷⁹ "[...] I do not understand the data-analysis 'yardsticks' which the authors proffer in place of classical tests of significance. In what probabilistic or other units are the yardsticks calibrated?" Ehrenberg in Brown/Durbin/Evans (1975, S. 177). Fisk weist darauf hin, daß mit einem solchen informellen Ansatz (Interpretation von Graphiken) die Gefahr von Fehlinterpretationen bestehe. So könne es durchaus möglich sein, daß die beobachteten Muster durch andere Ursachen als die vermuteten (nichtkonstante Parameter) hervorgerufen werden. Vgl. Fisk in Brown/Durbin/Evans (1975, S. 164). Andererseits ist eine solche Kritik auch hinsichtlich formaler Tests berechtigt.

³⁸⁰ Vgl. Harvey in Brown/Durbin/Evans (1975, S. 180).

The Bayesian approach offers a more formal framework for the inclusion of personal judgements [...].³⁸¹

Einen entscheidenden Punkt sprach schließlich Young an. Er wies darauf hin, daß die von den Autoren vorgestellten Techniken lediglich diagnostische Hilfen seien, die aber keine Antwort auf die Frage erlaubten, wie denn im Falle von nichtkonstanten Parametern weiter vorzugehen sei.³⁸² Die Antwort, die auch von mehreren anderen Diskussionsteilnehmern als Alternative genannt wurde, war der Kalmanfilter. Brown, Durbin und Evans verwiesen in diesem Zusammenhang auf die ingenieurwissenschaftliche Literatur.³⁸³

Im Rahmen der Statistik hat es seit den siebziger Jahren zwar eine Fülle von Untersuchungen über Modelle mit variablen Parametern gegeben,³⁸⁴ doch kein Verfahren ist bislang so allgemein akzeptiert wie die Testprozeduren von Brown, Durbin und Evans, so daß wir uns hier lediglich auf einen Hinweis beschränken wollen.

Geht man in einem allgemeinen linearen Modell,

$$(14) \quad y_t = \theta X_t + \varepsilon_t$$

in dem θ ein Parametervektor und X eine Regressormatrix darstellt,³⁸⁵ von Parametern aus, die sich im Zeitablauf verändern, könnte man für diese Änderung wiederum ein stochastisches Gesetz unterstellen. Dann erweitert sich das vorangegangene Modell z. B. zu

$$(15) \quad \begin{aligned} y_t &= \theta X_t + \varepsilon_t \\ \theta_t &= \theta G_{t-1} + \xi_t, \end{aligned}$$

wobei die Parameterentwicklung über einen bestimmten Algorithmus, eben den Kalmanfilter-Algorithmus, geschätzt werden kann. Die erste Gleichung wird als Meßgleichung, die zweite als Zustandsgleichung bezeichnet. In ingenieurwissenschaftlichen Zusammenhängen ist der Zustand oftmals die Position und Geschwindigkeit eines Körpers. In solchen physikalischen Systemen sind eine Reihe von Parametern des obigen Modells durch physikalische Gesetze vorgegeben. In praktischen Anwendungen, wie z. B. in der Ökonomie, sind mit einer solchen Modellierung allerdings zahlreiche Probleme verbunden, da keine Theorie Parameterwerte vorgibt und daher alle Parameter aus den Daten geschätzt werden müssen.³⁸⁶ In solchen Fällen behilft man sich zumeist mit Ad-hoc-Annahmen.³⁸⁷

³⁸¹ Smith in Brown/Durbin/Evans (1975, S. 176).

³⁸² Vgl. Young in Brown/Durbin/Evans (1975, S. 170).

³⁸³ Die gegenseitige Beeinflussung von Statistik und Ingenieurwissenschaften ist nach unserer Kenntnis noch nicht systematisch erforscht worden. Der Kalmanfilter-Algorithmus wurde erstmals von Kalman (1960) vorgestellt und zählt seither in zahlreichen technischen Bereichen zu einem unverzichtbaren Hilfsmittel: So diente dieser Algorithmus unter anderem zur Steuerung von Mondlandefähren.

³⁸⁴ Siehe dazu die Bibliographie von Hackl/Westlund (1989).

³⁸⁵ In dieser allgemeinen Form sind auch die Modelle (9), (10) und (11) als Spezialfälle enthalten.

³⁸⁶ Siehe den Kommentar von Davis zu West/Harrison/Migon (1985, S. 84).

c. Deskription und Inferenz in der explorativen Datenanalyse

In den letzten Jahren hat sich neben den Begriffen Deskription und Inferenz als dritte Säule der Statistik der Ausdruck "Datenanalyse", speziell die "explorative" Datenanalyse (EDA) etabliert. Der Ausdruck wurde von J. W. Tukey als Bezeichnung für eine ganz bestimmte Datenanalysekonzption geprägt und ist seither in einer ganzen Reihe von Bedeutungen verwendet worden. Während sich Hoaglin (1985) auf einen Überblick über die in diesem Rahmen angewandten Verfahren beschränkt, Polaseck (1994) in ihr eine moderne Bezeichnung für eine deskriptive Statistik sieht und Heiler/Michels (1994) den hypothesengenerierenden Charakter in den Vordergrund stellen, hat Good gar von einer "Philosophie" der explorativen Datenanalyse gesprochen. Für ihn ist die explorative Datenanalyse "[...] often informal, and perhaps EDA is more an art, or even a bag of tricks, than a science."³⁸⁷ Hypothesenformulierung sei ein elementarer Bestandteil des Ansatzes.

In der deutschen Forschung findet die explorative Datenanalyse weniger im Rahmen der theoretischen Statistik (noch weniger im Rahmen der Ökonometrie) als vielmehr in der empirischen Sozialforschung Beachtung. Vor allem Schnell (1994) hat sich dieser Thematik im Zusammenhang mit graphischen Datenanalysetechniken angenommen. Nach Schnell hat im allgemeinen eine Fehlrezeption der explorativen Datenanalyse stattgefunden, sei die eigentliche Intention Tukeys zum großen Teil verkannt worden.³⁸⁹ Er betont den iterativen Charakter des Datenanalyseprozesses, in dem explorative und konfirmatorische Analysen eine untrennbare Einheit bilden.³⁹⁰ Seiner Ansicht nach stehen bei Tukey keine "Parameterschätzungen oder gar Signifikanztests [...] im Mittelpunkt, sondern letztlich inhaltliche Modelle und die Abweichungen der eigenen Daten von diesem Modell."³⁹¹

Wie auch immer die explorative Datenanalyse im einzelnen interpretiert wird, sie beinhaltet stets den Versuch, einer Modellbildung, die nicht unabhängig von den Daten abläuft, Rechnung zu tragen. Insofern sind solche Ansätze für unser Thema sicher von Belang. Wir wollen im folgenden die Ansätze von J. W. Tukey und G. E. P. Box kurz vorstellen und uns insbesondere der Frage zuwenden, welche Rolle hierbei die beiden Erkenntnisziele Deskription und Inferenz spielen, zumal der letzte Punkt in der Diskussion offensichtlich eine untergeordnete Rolle zu spielen scheint.

³⁸⁷ Wir wollen hier nicht auf die komplizierten technischen Details eingehen. Eine Illustration der Berechnung von variablen Parametern findet sich bei Ledolter (1989). Aus bayesianischer Perspektive haben sich vor allem West/Harrison/Migon (1985), West/Harrison (1989) sowie Pole/West/Harrison (1994) mit diesen Modellen beschäftigt.

³⁸⁸ Good (1983b, S. 283).

³⁸⁹ Vgl. Schnell (1994, S. 328ff).

³⁹⁰ In diesem Zusammenhang beruft er sich vor allem auf G. E. P. Box.

³⁹¹ Schnell (1994, S. 327f). Dabei stellt sich natürlich die Frage, was mit "inhaltlich" gemeint ist. Da es sich hier um Statistik/Datenanalyse handelt, können damit nur statistische Modelle (d. h. systematische Komponenten) gemeint sein. Die von ihm mehrfach als bedeutende Aufgabe der Datenanalyse angesehene Residuenanalyse ist aber wohl schwerlich unabhängig von Parameterschätzungen denkbar.

Die Konzeption von J. W. Tukey

Die explorative Datenanalyse wird heutzutage zumeist mit dem Namen J. W. Tukeys, vor allem mit seinem Lehrbuch von 1977 verbunden.³⁹² Tukey ist für uns von besonderem Interesse, weil sich bei ihm umfangreiche Ausführungen zu "Philosophie und Prinzipien" der Datenanalyse finden. Wir wollen uns im folgenden auf wesentliche Aspekte beschränken.³⁹³ Dabei werden wir seine Position anhand seines vielleicht einflußreichsten Beitrags darstellen und einige Punkte, die für uns von Interesse sind, gegebenenfalls ergänzen.

Als bahnbrechender Beitrag wird in der Statistik allgemein sein 1962 in den *Annals of Mathematical Statistics* veröffentlichtes 48-Punkte-Programm zur "Zukunft der Datenanalyse" gesehen.³⁹⁴ Diese Arbeit, in der Tukey zum Teil bereits an anderer Stelle vorgetragene Aspekte systematisiert, zum Teil neue Fragen aufwirft, ist in mehrfacher Hinsicht bemerkenswert.³⁹⁵ Im Gegensatz zu anderen Artikeln der Zeitschrift enthält dieser Beitrag keine formalen Ableitungen und Beweise, propagierte keine Optimalitätskriterien. Statt dessen stellte er in 48 durchnummerierten Abschnitten fest, daß bei der Analyse von "real-world data" andere Probleme relevant sind. Tukey legte Wert darauf, "Mathematiker" und "Statistiker" zu trennen, wobei er unter letzteren, sich eingeschlossen, "Datenanalytiker" verstand.

³⁹² Tukey (1977). John Wilder Tukey [1915-] genoß keine formale Schulausbildung, sondern wurde statt dessen zum überwiegenden Teil von seinen Eltern unterrichtet. Nach einem 1937 an der Brown University abgeschlossenen Studium der Chemie promovierte er 1939 in Princeton in Mathematik und blieb dort als *instructor* bis 1942; anschließend arbeitete er für das *Fire Control Research Office* in Princeton. 1945 kehrte er auf Initiative von S. S. Wilks an die mathematische Fakultät der Universität Princeton zurück und wurde dort Mitglied des *statistics department*. Im gleichen Jahr wurde er Mitglied des Forschungsteams der Bell Laboratories. Beide Positionen hatte er bis zu seiner Pensionierung 1985 inne. 1957/58 war er Fellow des *Center for Advanced Study in Behavioral Sciences*. Seit dieser Zeit hat er sich auch immer wieder mit der Anwendung statistischer Methoden in den Verhaltens- und Sozialwissenschaften auseinandergesetzt. Sein Hauptaugenmerk galt jedoch stets den "harten" Wissenschaften.

³⁹³ Tukey hat in einem Zeitraum von über fünfzig Jahren mehr als 800 Arbeiten publiziert. Ein Teil seiner Publikationen ist zusammen mit einer Reihe von bis dahin unveröffentlichten Artikeln in der bislang achtbändigen Ausgabe *The Collected Works of John W. Tukey* erschienen. Für unseren Zusammenhang sind vor allem die Bände von Brillinger (1984), Jones (1986) und Cleveland (1988a) heranzuziehen.

³⁹⁴ Tukey (1962). Jones (1992, S. 404) stellt rückblickend fest: "In perspective, this paper presented a clear challenge to established standards of academic statistics, and it outlined the framework for the future development of exploratory data analysis." Diese Betonung der Bedeutung hatte allerdings nicht zur Folge, daß die Kritik von Tukey zu einem allgemeinen Umdenken oder gar "Paradigmenwechsel" geführt hätte; schon gar nicht in den von Tukey vorrangig angesprochenen "harten" Wissenschaften. Vgl. zum folgenden auch Rahlf (1996a, S. 144f).

³⁹⁵ Das Manuskript ging bei der Redaktion der *Annals of Mathematical Statistics* (AMS) am 1. Juli 1961 ein, demselben Tag, an dem ein neuer Herausgeber (J. L. Hodges) sein Amt übernahm. Es erschien als Leitartikel der ersten Ausgabe von 1962, während eine Reihe von Beiträgen, die vom vorhergehenden Herausgeber (W. Kruskal) bereits zur Publikation vorgesehen waren, erst Ende 1962 erschienen. Der Beitrag umfaßt insgesamt 67 Seiten. Von den 147 Artikeln, die 1962 in den AMS publiziert wurden, haben 25 % bis zu 6 Seiten, 50 % bis zu 8 Seiten und 75 % bis zu 14.5 Seiten Umfang. Der zweitlängste Beitrag umfaßt 29 Seiten. Vgl. Jones (1992).

Der Beitrag beginnt nicht mit einem Satz wie "Let X be a random variable...",³⁹⁶ sondern mit dem folgenden Statement:

"For a long time I have thought I was a statistician, interested in inferences from the particular to the general. But as I have watched mathematical statistics evolve, I have had cause to wonder and to doubt. [...] All in all, I have come to feel that my central interest is in data analysis, which I take to include, among other things: procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data."³⁹⁷

Er forderte statt dessen neue Fragen, neue Antworten in "unfamiliar summaries", insgesamt "more realistic frameworks", die die Existenz von "spotty data" berücksichtigen. Gerade in den Sozialwissenschaften genießt Tukey heute als Begründer der mittlerweile sehr beliebten "Explorativen Datenanalyse" hohes Ansehen. Eine genaue Lektüre seiner Texte kann dabei einige Fehlinterpretationen beseitigen:³⁹⁸ Tukey redete hier nicht etwa nichtparametrischen oder verteilungsfreien Methoden das Wort. Diese sollten sich beschränken "(i) upon situations where relative ranks are really all the data available, and (ii) upon situations where unusually quick or portable procedures are needed. In other situations it will be possible, and often desirable, to analyze the data more thoroughly and effectively by other methods."³⁹⁹ Die Absage an den allgemeinen Gebrauch der Normalverteilung bedeutete für ihn lediglich, diese durch eine andere, angemessenere Verteilung zu ersetzen. Die Konstruktion von Konfidenzintervallen und eine Forderung wie Effizienz gehörten ebenso zu seiner Konzeption von "Data Analysis" wie die "Robustness" und Monte-Carlo-Simulationen.⁴⁰⁰ Andererseits forderte er aber auch, "to give up the vain hope that data analysis can be founded upon a logico-deductive system [...] and to face up to the fact that *data analysis is intrinsically an empirical science*."⁴⁰¹ Insbesondere die Frage nach der Rolle von Modellen, mithin dem Teil, der gewöhnlich als gegeben angesehen wird, stellt sich in diesem

³⁹⁶ Das ist z. B. der Beginn der in Abschnitt 3 erwähnten Arbeit von Robbins (1955).

³⁹⁷ Tukey (1962 [1992, S. 408]).

³⁹⁸ Dies gilt selbst bei einer Beschränkung auf die "philosophischen" Schriften. Tukey (1986a, S. xl) betonte, daß bei einer alleinigen Betrachtung dieser Schriften der exploratorische Aspekt zu sehr im Vordergrund stehen könnte. Dies sei nicht sein Anliegen.

³⁹⁹ Tukey (1962 [1992, S. 411]).

⁴⁰⁰ Vgl. ebda., S. 418, 421, 446. Nach Lehmann (1992, S. 70) besteht für Tukey der Unterschied zwischen EDA (Exploratory Data Analysis) und CDA (Confirmatory Data Analysis) darin, die Daten in zwei Teile aufzuteilen: Anhand des ersten Teils sind die Modelle zu konstruieren, anhand des zweiten Teils diese Modelle mit der klassischen Inferenzstatistik zu evaluieren. In Tukey (1972, S. 760f) setzte sich die Datenanalyse aus drei Schritten zusammen: "The first stage is exploratory data analysis, which does not need probability, significance, or confidence, and which, when there is much data, may need to handle only either a portion or a sample of what is available. [...] The second stage is probabilistic." Schließlich könne noch, falls dieser zweite Schritt nicht zu überzeugenden Ergebnissen führe, ein dritter Schritt des "mustering and borrowing strength" hinzugefügt werden. Siehe auch Tukey (1980a).

⁴⁰¹ Tukey (1962 [1992, S. 448ff]). Hervorhebung im Original. Ähnlich Tukey (1986a, S. xxxix): "The philosophy [...] is far more based on a 'bottom up' approach than on a 'top down' one".

Zusammenhang. Tukey kritisierte hier die übliche Vorgehensweise, die von “gegebenen” Modellen ausging:

“Most textbook discussions of statistics or data analysis take the model (in all its aspects) as given – almost as handed down from above, as the tablets of stone were to Moses. This may have become a custom because mathematical theory could be developed starting from unquestioned assumptions and such theory was easier to teach than data analysis.”⁴⁰²

Für ihn ist die empirische Forschung im Rahmen der Schritte “conjecture-design-experiment-analysis” stets ein iterativer Prozeß, der sich nicht linear, sondern zyklisch vollzieht.⁴⁰³ Bei diesem Vorgehen werden theoretische Erklärungen immer wieder revidiert und erklären Phänomene des vorherigen Zyklus. Ausgangspunkt seien in der Regel empirische Regularitäten, die langlebiger als Theorien zu ihrer Erklärung seien und in der Regel den Ausgangspunkt für eine Forschung darstellen.⁴⁰⁴ Tukey hat 1984 seine Sicht in dem folgenden Schaubild zusammengefaßt.

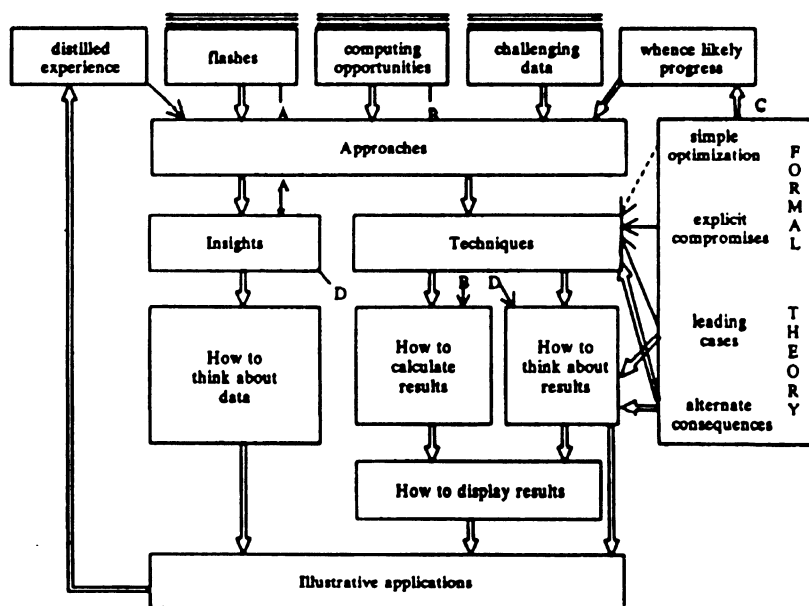


Abb. 4: Struktur der Datenanalyse nach Tukey. Quelle: Tukey (1984, S. xli).

⁴⁰² Mallows/Tukey (1982 [1986, S. 893f]).

⁴⁰³ U. a. in Tukey (1986b, S. 223f) berief er sich – im Zusammenhang experimenteller Forschung – explizit auf den frühen G. E. P. Box.

⁴⁰⁴ Vgl. ebda., S. 224.

Wie sind solche Aussagen in das bestehende Schema der statistischen (Inferenz-) Theorie einzuordnen? Es drängt sich die Frage auf, weshalb sich Tukey trotz seiner Betonung einer induktiven Vorgehensweise nicht etwa explizit zu einer bayesianischen Konzeption bekennt. Außer in der gemeinsam mit F. Mosteller verfaßten Arbeit⁴⁰⁵ finden wir lediglich in Tukey (1972) eine kurze explizite Behandlung des Themas. Diese beginnt mit den folgenden Worten: "On the last day of August, 1971 [dem Tag, an dem Tukey diesen Text öffentlich vortrug, T. R.], the place of Bayesian inference is to me a puzzle, partly because I am not sure what is the essence of Bayes."⁴⁰⁶

Seiner Ansicht nach besteht die Essenz des bayesschen Ansatzes aus vier Elementen:

1. In die Analyse sollte unser gesamtes Wissen einbezogen werden, nicht nur das durch die Daten gegebene,
2. man sollte nur die kombinierte Evidenz aus A-priori-Wissen und der Evidenz der Daten betrachten,
3. diese Kombination sollte über das Bayes-Theorem vollzogen werden, in der beide Formen von Wissen in ein Modell mit wenigen Parametern eingehen,
4. alle Techniken sollten auf diesem Theorem aufbauen.⁴⁰⁷

Während Tukey den ersten Punkt akzeptieren konnte, war er gegenüber dem zweiten skeptisch, "especially since the discovery of the irrelevance of past knowledge to the data before us can be one of the great triumphs of science."⁴⁰⁸

Der dritte Punkt sei nur zu akzeptieren, wenn das gesamte (Vor-) Wissen in eine sparsame parametrische Form gekleidet werden könne (dann sei die bayessche Lösung allerdings die beste), und der vierte Punkt schließlich sei auch zu befürworten, wenn es sich bei dem Vorgehen um solch überzeugende Arbeiten handle, wie die Untersuchung von Mosteller/Wallace (1964) über die Autorenschaft der Federalist Papers, wobei er allerdings fragt, ob gleichwertige Ergebnisse nicht auch mit klassischen nichtbayesianischen Ansätzen hätten erreicht werden können, wenn nur Wissenschaftler vom Format der beiden Autoren dies versucht hätten. Tukey resümierte:

"Bayes is still a puzzle to me; I hope each of the varied leaders in the use of Bayesian methods will gradually become willing to put his or her own essence of Bayes in simple terms for all to ponder upon. Clearly there are places to use Bayes and places to shun it. We do not have an adequate understanding of which is which."⁴⁰⁹

⁴⁰⁵ In Mosteller/Tukey (1968, S. 160-183) wurden bayesianische Elemente zwar relativ ausführlich behandelt, vermutlich aber eher durch Mosteller.

⁴⁰⁶ Tukey (1972 [1986, S. 770]).

⁴⁰⁷ Vgl. ebda., S. 771.

⁴⁰⁸ Ebda., S. 771. Das Argument ist allerdings für sequentielle bayessche Verfahren unzutreffend.

⁴⁰⁹ Ebda., S. 772. Hervorhebung im Original. Diese Position vertrat er im wesentlichen auch noch in Tukey (1986a, S. xlf).

Wir wollen die Behandlung des Tukeyschen Ansatzes mit einer kurzen Erwähnung des Bereiches der Zeitreihenanalyse und der Bedeutung statistischer Graphiken abschließen.

Nach eigenen Aussagen hat mehr als jedes andere Gebiet der Statistik die Spektralanalyse seine Datenanalysekonzeption beeinflusst.⁴¹⁰ Zwar hat Tukey auf diesem Gebiet entscheidende Beiträge geleistet, doch hat seine Datenanalysekonzeption die Entwicklung der Zeitreihenanalyse nicht beeinflusst. Über die Gründe für diese ausgebliebene Rezeption kann nur spekuliert werden. Nicht ohne Bedeutung ist sicher erstens seine Ablehnung der sich großer Popularität erfreuenden ARIMA-Modelle gewesen,⁴¹¹ hinzu kam zweitens die Orientierung seines Interesses auf Bereiche, die nicht ohne weiteres auf die Probleme anderer Wissenschaften übertragbar waren (v. a. aus dem geophysikalischen Bereich), und drittens verwendete er gerade in diesem Bereich eine Sprache, die dem Leser eine (zu) große Portion Umdenkens abverlangt.⁴¹²

Der letzte hier zu behandelnde Punkt ist die Bedeutung statistischer Graphiken im Rahmen einer statistischen Methodologie. Auch auf diesem Gebiet gilt Tukey als Pionier, und nach dem oben kurz angerissenen Ansatz ist es nur folgerichtig, daß graphischen Darstellungen darin eine grundlegende Bedeutung zukommt.⁴¹³ Dieser Aspekt wurde vor allem von seinem Schüler W. S. Cleveland ausgebaut.⁴¹⁴

Graphische Datenanalyse ist für Cleveland wie für Tukey nur ein Teilbereich. Mindestens ebenso wichtig ist die Datenkommunikation: "when the analyst wants to communicate data to others".⁴¹⁵ Entscheidend bei der Konstruktion einer Graphik sei eine möglichst effiziente und präzise Dekodierung, die auf einer wissenschaftlichen Grundlage basieren sollte: "It is only through scientific study of visual perception that informed judgments can be made about display methods. The display methods of elements rest on a foundation of scientific enquiry."⁴¹⁶

⁴¹⁰ Vgl. Tukey (1984, S. xxxix).

⁴¹¹ Er sah den Anwendungsbereich solcher Modelle als sehr beschränkt an: "The [...] category [der ARIMA-Modelle, T. R.] arises when the combination of amount of data with distinctiveness of behavior is weak enough for us to do no more than ask: 'By what simple mechanism could the data we see have come from a white-noise process [...]?'". Mallows/Tukey (1982 [1986, S. 929]). Das Ziel einer solchen Analyse könne nur in einer Deskription oder Prognose bestehen.

⁴¹² Dies gilt vor allem für die nicht-technischen, konzeptionellen Beiträge. Siehe hier vor allem Tukey (1980b).

⁴¹³ Ein Blick in das Inhaltsverzeichnis von Tukey (1977) zeigt dies vielleicht am deutlichsten. Vgl. auch Cleveland (1988a) sowie zur Einführung Cleveland (1988b).

⁴¹⁴ W. S. Cleveland hat bereits unzählige Beiträge zu diesem Themenkreis veröffentlicht, unter anderem in so bedeutenden Zeitschriften wie dem *Journal of the American Statistical Association* oder dem *Journal of the Royal Statistical Society*, manchmal unter kuriosen Titeln, wie etwa "Variables on Scatterplots Look More Highly Correlated when the Scales are Increased". Eine Synthese findet sich in Cleveland (1993) und Cleveland (1994). In "Elements of Graphing Data", dessen erste Auflage 1985 erschien, liegt der Schwerpunkt auf allgemeinen Konstruktionsprinzipien und Aspekten der Wahrnehmung. Das Buch "Visualizing Data" ist dagegen vorrangig graphischen Methoden und Hilfsmitteln gewidmet. Vgl. zum folgenden Rahlf (1995a).

⁴¹⁵ Cleveland (1994, S. 1).

⁴¹⁶ Ebda.

Welche Stellung hat nach diesem Ansatz der Bereich der graphischen Datenanalyse in der allgemeinen statistischen Methodologie? Für Cleveland ist dabei insbesondere das Verhältnis zur Inferenzstatistik von Bedeutung.⁴¹⁷ Visualisierung ermöglicht für ihn die Überprüfung von Verteilungen und Wahl der geeigneten Transformation, um damit eine Voraussetzung für die Anwendung inferenzstatistischer Aussagen zu schaffen. Er schließt einen hypothesengenerierenden Aspekt graphischer Darstellungen nicht aus und warnt vor einer "rote data analysis" (S. 85). Er lehnt aber keinesfalls Signifikanztests ab. Auch die drastische Bewertung einer Studie, die eine Hypothese testet, bedeutet keine generelle Ablehnung, sondern nur eine Forderung nach der richtigen Anwendung, d. h. nach einer angemessenen Modellierung: "For *this* analysis [gemeint ist die angemessene Modellierung, T. R.], we can venture the standard methods to make probabilistic inferences" (S. 178). Die graphischen Darstellungen dienen zwar dazu, geeignete Transformationen und Modelle zu finden; ihre letztendliche Überprüfung könnten sie jedoch nicht ersetzen.

In Deutschland wird eine graphisch gestützte Datenanalyse als Konzept vor allem von Schnell (1994) propagiert. Schnell widmet ihrer Einordnung in den Rahmen der Statistik ein ganzes Kapitel, in dem er in zum Teil polemischer Form einen Großangriff auf die klassische Inferenzstatistik und die seiner Meinung nach falsche Rezeption der explorativen Datenanalyse startet. Danach sollte die (explorative) Anwendung von Graphiken in einem "iterativen" Prozeß zu einer angemessenen Modellbildung führen. Insbesondere diene die Graphik hierbei nicht nur der Hypothesenüberprüfung und -modifizierung, sondern auch der Hypothesengenerierung (S. 331). Seiner Ansicht nach ist "[...] die Debatte um die weitgehende Nutzlosigkeit der Signifikanztests im Rahmen der tatsächlichen Forschung nahezu vollständig vergessen bzw. gar nicht erst rezipiert worden" (S. 338). Er glaubt, "Signifikanztests beantworten daher eine Frage, die kaum interessiert" (S. 339). Sicher ist die von ihm angegriffene kritiklose, mechanische Anwendung von Signifikanztests in vielen Fachveröffentlichungen problematisch – wenn etwa Voraussetzungen des angewandten Tests nicht gegeben sind. Das ändert aber nichts an der grundsätzlichen Notwendigkeit dieser Tests in Situationen, in denen eine Stichprobenkonzeption sinnvoll ist. Diese Haltung läßt sich jedenfalls nicht mit den von ihm angeführten Konzeptionen von Tukey oder Box rechtfertigen.

Die Konzeption von G. E. P. Box

Anders als J. W. Tukey hat sich G. E. P. Box in seiner Datenanalysekonzeption auch für die Verwendung bayesianischer Elemente ausgesprochen.⁴¹⁸ Er hat seinen Ansatz bereits in den fünfziger Jahren, damals ausschließlich bezogen auf experimentelle Designs in den Natur- und Ingenieurwissenschaften, entwickelt. Man wird Box vermutlich am ehesten gerecht, wenn man ihn als "Instrumentalisten"

⁴¹⁷ Er schließt in *Visualizing Data* jedes Kapitel mit einem Abschnitt "Visualization and Probabilistic Inference" ab.

⁴¹⁸ Siehe zum folgenden vor allem Box (1976), Box (1979) sowie Box (1980).

bezeichnet. Am deutlichsten wird diese Grundeinstellung bezüglich der Bedeutung von Modellen: "Models, of course, are never true, but fortunately it is only necessary that they be useful."⁴¹⁹ Der Prozeß der Modellbildung, der in den Lehrbüchern der Statistik in der Regel nicht berücksichtigt werde, sei von zentraler Bedeutung. Dieser informelle, zumeist graphische Techniken verwendende Teil der Datenanalyse sei bislang jedoch eher wie schwarze Magie behandelt worden.⁴²⁰

Seiner Ansicht nach ist es daher völlig unmöglich, ein "richtiges" Modell aufzustellen:

"Since all models are wrong the scientist obtains a 'correct' one by excessive elaboration. On the contrary following William of Occam⁴²¹ he should seek an economical description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist so overelaboration and overparameterization is often the mark of mediocrity. [...] Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad. [...] the statistician knows, for example, that in nature there never was a normal distribution, there never was a straight line, yet with normal and linear assumptions, known to be false, he can often derive results which match, to a useful approximation, those found in the real world."⁴²²

Die wichtigste Aufgabe des Statistikers sah Box ähnlich wie Tukey nicht in der Entwicklung mathematisch ausgefeilter Techniken, sondern in der Analyse konkreter Daten. Entscheidend war in diesem Zusammenhang für ihn die Iteration zwischen Theorie (Hypothesen, Modelle) und Praxis (Daten, Fakten). Ein aufgrund von vorhandenen Daten aufgestelltes Modell werde einer kritischen Prüfung unterzogen, möglicherweise durch neue Fakten revidiert, wiederum geprüft usw. Auf jeden Fall sollte man von einem tatsächlichen Problem ausgehen und nicht von der Lösung mathematischer Spezialfragen.⁴²³

Um die Parameter eines Modells zu schätzen, bevorzugte Box einen bayesianischen Ansatz.⁴²⁴ Die damit verbundene Notwendigkeit der Spezifikation einer den "prior belief" ausdrückenden A-priori-Verteilung ist seiner Ansicht nach kein bayesianisches Spezifikum, denn: auch eine Modellannahme im Rahmen des klassischen Ansatzes, wie z. B. die Normalverteilung der Fehler, sei nichts anderes als ein "prior belief":

"I believe that it is impossible logically to distinguish between model assumptions and the prior distribution of the parameters. The model *is* the prior in the wide sense that it is a probability

⁴¹⁹ Box (1979, S. 2).

⁴²⁰ Vgl. ebda., S. 3.

⁴²¹ Wilhelm von Ockham [ca. 1285-1347], englischer Theologe und Philosoph, war Haupt einer Schule, die Glauben und Wissen bzw. Theologie und Philosophie trennen wollte.

⁴²² Box (1976, S. 792).

⁴²³ Vgl. ebda., S. 798. Box sah sich in dieser Argumentation in einer Linie mit R. A. Fisher.

⁴²⁴ Box' Interesse für bayesianische Konzepte entstand aus Überlegungen bezüglich des Problems der Robustheit von Schätzungen und Tests. "[...] I really couldn't imagine that all this tremendous complication that you got into with sampling theory could possibly be necessary. I mean there must be some better way of thinking about it without getting into all that rigamalore." Box in DeGroot (1987, S. 251).

statement of all the assumptions currently to be tentatively entertained *a priori*. On this view traditional sampling theory was of course not free from assumptions of prior knowledge.”⁴²⁵

Auf der anderen Seite sollte jedoch die *Modellkritik* nicht in einem bayesianischen Kontext, sondern auf informelle Art⁴²⁶ sowie durch herkömmliche Signifikanztests geschehen:⁴²⁷

“It seems to me that the significance test idea is natural and indeed a necessary part of the conduct and management of every day life, and I find it hard to understand the horror with which it is sometimes greeted nowadays. The process of modification of belief occurs in two stages: (a) the recognition that the data do not fit with the presently entertained model of the world, (b) the later consideration of what are alternative models that might better explain the data.”⁴²⁸

Er führte ein aufschlußreiches Beispiel an:

“For example, suppose I have an office that looks onto, say Oxford Street in London, normally thronged with people. One day I look out of the windows at 11 o'clock in the morning and notice that there are only two people in the whole street. My initial reaction surely is that on the null (status quo) model this is an unusual event possible worthy of further investigation. *Alternative models that might explain the phenomena come later*. These might posit that the street has been blocked off for a ceremonial occasion, that there is a bomb scare, or that it is a Sunday, etc. *But notice that the basis of the initial reaction, which requires no alternatives*, is surely that I have (or could have) looked out of the window on many previous occasions and rarely have I (would I have) seen as few as two people in the street. The motivation is economy of effort and is employed by all of us hundreds of times in our daily lives - *when the null model is plausible I will not worry, but when data make it implausible perhaps I should be concerned*.”⁴²⁹

Mit dieser Signifikanztestkonzeption sprach Box sich eindeutig für die Position Fishers und gegen die Neyman-Pearson-Theorie aus, die stets eine Alternativhypothese fordert.⁴³⁰

Alles in allem läßt sich sowohl die Position von Box als auch diejenige Tukeys nicht einer bestimmten Richtung zuordnen. Beiden ist gemein, daß sie die herkömmliche Vorgehensweise als unzureichend betrachteten, weil sie der “Realität”, insbesondere der Interaktion zwischen Theorien und Daten nicht gerecht werde. Sie forderten daher beide einen informellen Umgang mit Daten, in dem graphische Techniken verwendet werden sollten. Zumindest ein Unterschied besteht jedoch: Während Tukey bayesianische Ansätze ablehnt, sind sie für Box die beste Möglichkeit, Parameter zu schätzen. Die Modellkritik sollte seiner Ansicht nach jedoch

⁴²⁵ Box (1980, S. 394). Hervorhebungen im Original.

⁴²⁶ Dazu gehöre vor allem die visuelle Inspektion von Residuenplots.

⁴²⁷ Siehe z. B. Box (1980, S. 384): “On this view [...] both processes [Stichproben- und bayesianische Inferenz, T. R.] would have essential roles in the continuing scientific iteration just as the two sexes are required for human reproduction.” Vgl. dazu auch Hodges (1990).

⁴²⁸ Box (1980, S. 429). Insofern mag es überraschen, wenn sich Schnell (1994) trotz mehrfacher vehementer Ablehnung von Signifikanztests (z. B. S. 328, 338ff, vgl. dagegen S. 62, 242f, 271ff) bei seiner Datenanalysekonzeption explizit auf Box beruft (S. 329).

⁴²⁹ Box (1980, S. 429). Hervorhebungen von uns.

⁴³⁰ Siehe oben, Abschnitte A 3 d, e.

im Rahmen einer Stichprobenkonzeption stattfinden. Beide Konzeptionen schließen sich demnach nicht aus, sondern ergänzen sich.

d. Ausblick: Neuere Entwicklungen

Bevor wir die Konzepte der Statistik zusammenfassend beurteilen, wollen wir noch einen Ausblick auf die Diskussion um den Prozeß der Modellkonstruktion werfen, die in den letzten Jahren in der Statistik zunehmend an Bedeutung gewann.

Als einer der ersten formulierte Michael C. Lovell 1983 die Konsequenzen des *data mining*.⁴³¹ Dieser Ausdruck, der seiner Ansicht nach in gewisser Weise sogar ein Forschungsparadigma bezeichnet, findet sich in der praktizierten Vorgehensweise der meisten statistischen Methoden anwendenden Untersuchungen wieder. Diese Vorgehensweise sei charakterisiert durch Formulierungen wie "aus Platzgründen werde nur das endgültige oder beste Modell präsentiert", "die signifikanten Variablen wurden in einem schrittweisen Regressionsverfahren ausgewählt", "das Modell wurde anhand der Daten bestimmt" usw. Solch ein Vorgehen sei in den letzten Jahren vor allem durch die rasante Entwicklung der Computertechnologie begünstigt worden. Vor allem die Verfügbarkeit einfach zu bedienender Computerprogramme habe es ermöglicht, anhand ein und desselben Datensatzes eine Vielzahl von Modellen zu testen. Diese Vorgehensweise werde auf der anderen Seite in den herkömmlichen Lehrbüchern nicht berücksichtigt. Die dort angegebenen Standardfehler von Schätzungen setzen stets a priori gegebene Modelle voraus. Die durch eine solche Vorgehensweise für einzelne Parameterschätzer möglicherweise zu engen Konfidenzintervalle seien hierbei noch das geringste Problem. Lovell veranschaulichte die möglichen Konsequenzen eines solchen Vorgehens anhand eines fiktiven Beispiels.

Angenommen, ein Forscher habe eine gleiche Anzahl von Variablen, die als mögliche erklärende Variablen in Betracht kommen. Da lediglich einfache Modelle betrachtet werden sollen,⁴³² werde davon ausgegangen, daß höchstens zwei erklärende Variablen in dem Modell enthalten sein sollten. Konkret werde als funktionale Form das lineare Regressionsmodell

$$(16) \quad Y_t = \beta_0 + \beta_1 X_{it} + \beta_2 X_{jt} + \varepsilon_t$$

unterstellt, wobei X_{it} und X_{jt} auf eine bestimmte Art aus der Menge aller vorhandenen erklärenden Variablen ausgewählt worden seien und ε_t eine normalverteilte Zufallsvariable mit endlicher Varianz darstelle. Wenn nun der Forscher anhand der üblichen t -Statistiken die Signifikanz derjenigen erklärenden Variablen teste, die schließlich in die Regressionsgleichung eingegangen seien, stelle sich die Frage, wie das wirkliche Signifikanzniveau des Tests ist:

⁴³¹ Lovell (1983).

⁴³² Dies ist eine übliche Forderung, die gelegentlich mit "Occam's Razor" assoziiert wird. Vgl. Lovell (1983, S. 2).

“That is, what is the probability that the t -statistics in the final regression obtained by data mining will appear significant at the 5 % level if the null hypothesis is true, there being no causal relationship so that in fact $Y_t = \beta_0 + \varepsilon_t$?”⁴³³

Geht man von keiner speziellen Variablenselektionsprozedur aus, sondern wählt lediglich die Variablen aus, die signifikant sind, so sind die entsprechenden Wahrscheinlichkeiten offensichtlich binomialverteilt. Geht man zusätzlich – der Einfachheit halber – davon aus, daß zum einen die Varianz des Zufallsterms bekannt, zum anderen die Regressoren unkorreliert seien, ergibt sich folgendes Modell: Die Wahrscheinlichkeit, daß (genau) keine der c Variablen signifikant ist, beträgt bei einem nominalen Signifikanzniveau α

$$(17) \qquad \binom{c}{0} (1-\alpha)^c,$$

die Wahrscheinlichkeit, daß genau eine Variable signifikant ist, berechnet sich aus

$$(18) \qquad \binom{c}{1} \alpha^1 \times (1-\alpha)^{c-1}.$$

Für genau zwei Variablen beträgt die Wahrscheinlichkeit

$$(19) \qquad 1 - (1-\alpha)^c - \binom{c}{1} \alpha \times (1-\alpha)^{c-1}.$$

Das *wahre* Signifikanzniveau α^* , die korrekte Wahrscheinlichkeit, keinen Regressorkandidaten als signifikant zu bezeichnen, ergibt sich damit für verschiedene Werte von c wie in Tab. 2 angegeben:

Tab. 2: Wahrscheinlichkeit, anhand von zufälligen Daten “signifikante” Resultate zu erhalten, wenn die zwei besten Regressoren ausgesucht werden. Quelle: Lovell (1983, S. 4).

	Anzahl in Frage kommender Variablen									
	$c=2$	$c=5$		$c=10$		$c=20$		$c=100$		
Nominales Signifikanzniveau α	5%	5%	1%	5%	1%	5%	1%	5%	1%	
Anzahl "signifikanter" Koeffizienten:										
0	.903	.774	.951	.599	.904	.358	.818	.006	.366	
1	.095	.204	.048	.315	.901	.377	.165	.031	.370	
2	.0025	.023	.001	.086	.004	.264	.017	.962	.264	
Wahres Signifikanzniveau α^*	5%	12%	2.5%	22.6%	4.9%	40.1%	9.6%	92.3%	39.5%	

⁴³³ Ebda.

Die linke Spalte zeigt zum Vergleich die Situation, die die Theorie vorsieht: Alle zur Verfügung stehenden Variablen werden in die Regressionsgleichung einbezogen, und der Fehler erster Art (die Wahrscheinlichkeit, die Nullhypothese – keine erklärende Variable – zurückzuweisen, wenn sie wahr ist) ist exakt 5 %. Wenn nun die Suche nach “erklärenden” Variablen auf mehr als zwei Größen ausgedehnt wird, verändert sich das Signifikanzniveau jedoch. In der zweiten Spalte werden zwei von fünf Regressoren ausgewählt. Dies reduziert die Wahrscheinlichkeit, korrekterweise keine signifikanten Regressoren festzustellen, auf

$$(20) \quad (1-0.12)^2 = 0.774$$

oder erhöht das Signifikanzniveau auf 12 %. Entsprechend handelt es sich bei einem nominalen Niveau von 1 % faktisch um ein 5 %-Niveau. Erhöht sich die Anzahl der potentiell in Frage kommenden Variablen auf 10, so sinkt die Wahrscheinlichkeit, korrekterweise die Nullhypothese nicht zu verwerfen, auf 59.9 %. Das wahre Signifikanzniveau der *t*-Tests steigt auf 23 %.⁴³⁴

Insofern muß eine eindeutige Warnung vor einem naiven “data mining” ausgesprochen werden. Legt man die nominalen Signifikanzniveaus zugrunde, kann es – in wiederholten Situationen – mit hohen Wahrscheinlichkeiten zu Fehlurteilen kommen.⁴³⁵

Schon diese unrealistische Situation – bekannte Varianz und orthogonale Regressoren – mahnt zur Vorsicht. Gehen wir nun von einer unbekannten, zu schätzenden Residualvarianz aus. In diesem Fall wird durch die Auswahl der besten Regressoren die Residualvarianz systematisch unterschätzt; damit ist eine Überschätzung der *t*-Werte verbunden. Weiterhin ist in der Realität von korrelierten Regressoren auszugehen, wodurch sich das einfache binomiale Modell nicht mehr anwenden läßt.⁴³⁶

Den Versuch einer Synthese des Problems der Modellbildung und seiner Bedeutung im Rahmen der Inferenz hat vor kurzem Chris Chatfield zur Diskussion gestellt.⁴³⁷ Chatfield bezeichnet den Bereich der klassischen (Inferenz-) Statistik als “interesting, but narrow” (S. 419). Sein Anliegen ist, den Begriff der statistischen Inferenz zu erweitern und den *gesamten* Modellkonstruktionsprozeß mit einzubeziehen. Dieser besteht seiner Ansicht nach aus vier Hauptkomponenten:

1. der Modellspezifikation,
2. der Modellanpassung/Parameterschätzung,
3. der Modellvalidierung und
4. der Kombination von Daten aus mehreren Quellen (z. B. Meta-Analyse).

⁴³⁴ Lovell (1983, S. 3) hat zwar auch eine “Daumenregel” als Lösung angeboten, da wir hier aber lediglich das Problem illustrieren wollen, gehen wir nicht näher darauf ein.

⁴³⁵ Dennoch wird ein solches Vorgehen in den derzeit gängigen Softwarepaketen im Rahmen verschiedener Prozeduren wie der schrittweisen Regression praktiziert.

⁴³⁶ Lovell hat auch das Problem des *data mining* in der Ökonometrie untersucht. Darauf kommen wir weiter unten, Abschnitt B 3 a, noch einmal zurück.

⁴³⁷ Chatfield (1995).

Chatfield sieht in Anlehnung an Box die Konstruktion eines Modells als iterativen Prozeß, der die Elemente der Problemformulierung, des interdisziplinären Austausches mit anderen Disziplinen sowie die Interpretation und Mitteilung der Resultate beinhalte. Seiner Ansicht nach besteht hier ein Forschungsdesiderat: "The more recent references give guidance on model selection methods for choosing a 'best' model from two or more prespecified models having different structures, but rather little help on model formulation in its widest sense – how do you choose the models to be considered? This is arguably the most important and most difficult aspect of model building and yet is the one where there is least help [...]." ⁴³⁸ Die Unsicherheit im Rahmen statistischer Inferenz bestehe auf drei Ebenen:

1. Unsicherheit über die Struktur des Modells, die verursacht sein kann durch
 - a) Fehlspezifikation (z. B. Nichtberücksichtigung von Variablen),
 - b) Spezifikation einer Klasse von Modellen, von der das wahre Modell ein (unbekannter) Spezialfall ist (z. B. ARIMA-Modelle in der Zeitreihenanalyse),
 - c) Auswahl aus zwei oder mehr Modellen mit völlig verschiedenen Strukturen,
2. Unsicherheit hinsichtlich der Schätzungen der Modellparameter unter der Annahme, daß die Struktur des Modells bekannt ist und
3. unerklärte Zufallsstörungen in den beobachteten Variablen, selbst wenn sowohl Modell als auch Parameter bekannt sind. ⁴³⁹

Die statistischen Theorien beschränkten sich bislang auf die Punkte 2. und 3. sowie auf 1. b) (z. B. durch Selektionskriterien wie AIC). Die Punkte 1. a) und c) seien dagegen kaum untersucht worden. Dies ist nach Ansicht Chatfields um so problematischer, als Fehler, die aus 1. resultieren, in der Regel größere Auswirkungen haben dürften als Fehler, die aus 2. resultieren, auf der anderen Seite durch leistungsfähige Computer eine Vielzahl von Modellen für einen Datensatz evaluiert werden können und somit die Möglichkeit eines zufällig passenden Modells steigt, für das schätztheoretische Resultate ohne Bedeutung sind: "Statisticians must stop pretending that model uncertainty does not exist and begin to find ways of coping with it." ⁴⁴⁰

Er führt zwei anschauliche Beispiele an: Der einfachen Schätzung eines Mittelwertes einer Normalverteilung (z. B. separat oder im Rahmen einer Residuenanalyse) geht in der Regel eine Überprüfung der Verteilung entweder formal oder durch eine visuelle Inspektion voraus. Nur wenn die Daten diese Prüfung überstehen, wird zur Schätzung übergegangen, unter Umständen nach einer Transformation, die zu einer Normalverteilung führt. Bei der Schätzung wird dieser vorhergehende Schritt ignoriert. Chatfield fragt zu Recht: "What effect does this have? I am

⁴³⁸ Ebda., S. 419.

⁴³⁹ Vgl. ebda., S. 421.

⁴⁴⁰ Ebda., S. 422.

not aware of any help in the literature on this question".⁴⁴¹ Andererseits ist jedoch festzuhalten, daß die hier angesprochenen Probleme gerade in jüngster Zeit verstärkt thematisiert werden.⁴⁴²

5. Zusammenfassung und kritische Würdigung

Bevor wir uns den Konzepten der Ökonometrie zuwenden, wollen wir die in diesem Teil besprochenen Konzepte der Statistik noch einmal zusammenfassend beurteilen. Die Konzepte wurden in die Komplexe Deskription und Inferenz aufgeteilt. Allerdings mußte festgestellt werden, daß die theoretische Statistik hinsichtlich methodischer Fragen der Deskription von Daten bislang kaum konzeptionelle Aussagen, etwa wann der durchschnittliche Abstand als Deskription der Streuung der mittleren quadratischen Abweichung vorzuziehen ist, bereitstellte. Hervorzuheben bleibt jedoch, daß das universell angewendete Kleinstquadratverfahren als Ausgleichsverfahren zunächst eine Deskription darstellt. Seine Anwendung im inferenzstatistischen Sinne ist keineswegs eine so selbstverständliche "Wahrheit", wie es in der Literatur suggeriert wird.⁴⁴³

Wie entwickelten sich dagegen *stochastische* Denkweisen? Bis zur Mitte des 17. Jahrhunderts hatte die Wahrscheinlichkeitstheorie eher Denksportcharakter; es handelte sich um reine Kombinatorik. Die Wahrscheinlichkeit, bei einem Wurf eine bestimmte Zahl mit einem Würfel oder eine Seite einer Münze zu erhalten bzw. eine bestimmte Karte zu ziehen, konnte ohne tiefgreifende philosophische Überlegungen über das Wesen der Wahrscheinlichkeit angegeben werden. Die Wahrscheinlichkeit, z. B. bei zehn Münzwürfen genau viermal "Zahl" und sechsmal "Wappen" zu erhalten, war dann durch rein mathematisch-kombinatorische Überlegungen zu bestimmen, denn bei einem Münzwurf-"Experiment" existiert ein vollständig spezifiziertes theoretisches Modell: Die Ereignisse sind unabhängig voneinander, somit ist die Summe der Ereignisse binomialverteilt, der Parameter beträgt $\pi = 0.5$.

Die Fragen, die zu dieser Zeit in einem sozioökonomischen Kontext auftraten, waren jedoch nur scheinbar gleicher Natur. Auch Größen wie das allgemeine Geschlechterverhältnis, die Lebenserwartung, die Rate der Kindersterblichkeit, der Anteil der waffenfähigen Bevölkerung usw. wurden als gesetzmäßig angesehen. Konnte man die Sicherheit der ermittelten Resultate abschätzen?

Für unseren Zusammenhang spielt die ebenso interessante wie schwierige Frage, ob sich in solchen Interessen die Ergründung der Existenz und des Wirkens Gottes

⁴⁴¹ Ebda.

⁴⁴² Siehe etwa den Sammelband von Dijkstra (1988) sowie die Arbeiten von Cox (1990), Faraway (1992), Ehrenberg/Bound (1993), Hahn/Meeker (1993), Wild (1994), Draper (1995), Kass/Raftery (1995), Laud/Ibrahim (1995) oder O'Hagan (1995) sowie die Monographie von Hjorth (1994).

⁴⁴³ Ehrenberg/Bound (1993, S. 188) sprechen in diesem Zusammenhang von dem vorherrschenden Glauben, daß "all men are endowed with certain inalienable rights, among them life, liberty, the pursuit of happiness, and least squares."

durch den Nachweis von Ordnung offenbarte oder bereits naturwissenschaftliche Motivationen im Vordergrund standen, keine Rolle.⁴⁴⁴ Wichtig ist jedoch, daß Untersuchungen wie die eines J. Graunt anhand der Totenregister von London oder eines E. Halley über die Geburten- und Todesfälle in Breslau die Aufmerksamkeit von Mathematikern wie G. W. Leibniz, J. Bernoulli oder A. de Moivre erregten und zum Nachdenken über das Problem der *Inferenz* zwangen. Der Anteil der Personen, die eine bestimmte Merkmalsausprägung aufwiesen, war unbekannt, solange man ihn nicht für eine gegebene (Teil-) Population auszählen konnte. Hier war man eventuell berechtigt, das Binomialmodell zu unterstellen, aber es existierte beileibe keine Theorie, die einen Wert für den Parameter postulierte, den es zu überprüfen galt. Vielmehr sollte dieser Wert eben erst anhand der Daten ermittelt und durch ein Intervall ein Maß für die Genauigkeit der "Schätzung" angegeben werden. Daher war eine *Inversion* von Bernoullis Theorem nötig, doch weder Bernoulli noch de Moivre konnten diesen Schritt vollziehen. Wir wollen uns hier der These Stiglers anschließen und davon ausgehen, daß die konzeptionellen Schwierigkeiten erst über den Umweg der Fehlerrechnung schließlich durch T. Bayes und P. S. Laplace überwunden werden konnten. Diese "kopernikanische Revolution"⁴⁴⁵ in der Entwicklung der theoretischen Statistik knüpfte an die Intention Bernoullis an. Daß diese Konzeption heute mit dem Namen von Bayes und nicht von Laplace verbunden wird, ist eher ein Kuriosum. Bayes hatte eine entscheidende Idee, die jedoch zur gleichen Zeit und vermutlich unabhängig davon auch Laplace entwickelte. Laplace hatte jedoch darüber hinaus eine systematische Wahrscheinlichkeitstheorie konstruiert.

Die Grundlegung der *sozialwissenschaftlichen Statistik* durch A. Quetelet entwickelte sich nicht, wie man denken könnte, aus den Wurzeln der Politischen Arithmetik, sondern aus der astronomischen und physikalischen Fehlerrechnung. Von entscheidender Bedeutung wurde der "Zentrale Grenzwertsatz." Laplace hatte die Normalverteilung zum universalen Gesetz erhoben: Wenn die "Fehler" zahlreich und unabhängig waren und sich gegenseitig aufhoben, war die resultierende Verteilung "normal".

Quetelet knüpfte mit den *Methoden* der Astronomie und Geodäsie an die *Interessen* der politischen Arithmetiker und Wahrscheinlichkeitstheoretiker des 17. Jahrhunderts an. Seine Suche nach Gesetzmäßigkeiten im Gesellschaftsleben stieß jedoch zunächst auf einen Widerstand, mit dem sich bereits Bernoulli konfrontiert sah und der nun unter der Bezeichnung "Keverberg-Argument" wiederauflebte. Man konnte nicht wie bei einem Münz- oder Würfelwurf oder wie im Rahmen der Fehlertheorie von konstanten, individuell unbedeutenden Störfaktoren ausgehen, sondern mußte "die Umstände" berücksichtigen. Diese Umstände waren etwa bei der Verteilung der Mortalität die Berufs- oder Ortszugehörigkeit, bei der Geburtenrate der Monat usw. Je nach Ausprägung dieser Umstände ergaben sich unterschiedliche Verteilungen. Sowohl die Auswahl der beeinflussenden Faktoren als auch die Anzahl der Kategorien war jedoch nicht theoretisch festzulegen.⁴⁴⁶

⁴⁴⁴ Vgl. dazu Schneider (1988, S. 180).

⁴⁴⁵ Stigler (1986, S. 122).

Während Quetelet die “Umstände” nur in Form von Kategorien (etwa die Mortalität in Abhängigkeit vom Geschlecht) betrachtete, ging F. Galton einen entscheidenden Schritt weiter. Auch in der Biologie existiert kein “wahrer” oder “fester” Parameterwert wie die Länge einer astronomischen Entfernung oder der Durchmesser eines Planeten. Hier mußte eine solche Vorstellung daher Befremden hervorrufen. Die exakte Höhe z. B. einer Pflanze wie Habichtskraut zu bestimmen, hat allenfalls theoretischen Wert, da “Habichtskraut” kein Unikum, sondern eine Gattungsbezeichnung ist. Das Hauptziel statistischer Erkenntnis verschob sich daher: Nicht mehr die Ermittlung von Wahrscheinlichkeiten von (diskreten) Ereignissen (Bernoulli, de Moivre), nicht mehr die Ermittlung einer konstanten Größe, wie der Erdumfang oder eine Planetenbahn, sondern die *Verteilungen selbst*, die Variationen standen nun im Mittelpunkt.⁴⁴⁷

Galton ging in einem zweiten Punkt entscheidend über Quetelet hinaus. In dem von ihm betrachteten biologischen Kontext lag es näher, von stetigen anstelle von kategorialen beeinflussenden Faktoren auszugehen. Wenn z. B. die menschliche Körpergröße normalverteilt ist, so muß dies sowohl für Eltern als auch für Kinder gelten. Die Größe der Eltern zeigt in seinen Daten jedoch einen Einfluß auf die Größe der Kinder, die Daten waren somit *gemeinsam* oder *bivariat* normalverteilt. In diesem Kontext entstand das Konzept der Korrelation: *Wenn* zwei Größen bivariat normalverteilt sind, *dann* ist der Zusammenhang zwischen ihnen linear und läßt sich durch einen Korrelationskoeffizienten ausdrücken.

Für Galton waren die “Umstände” also Teil eines gemeinsamen Phänomens. Hiermit stellte sich nun die Frage, welche Größen als entscheidend angesehen werden sollten und welche nicht. Wenn die Realisation eines Ereignisses als *zufällig* angesehen werden kann, das heißt alle Bedingungen, die die konkrete Realisation eines Zufallsereignisses determinieren, zusammengefaßt werden und man sich nur für die Gesamtheit der Einflüsse für alle Realisationen interessiert, dann ist die “Erhebungsmodalität” (das Wann und Wo) ohne Bedeutung. Dennoch würde man beispielsweise zögern, mit der auf einer Wiese, an einem Ort, zu einer Zeit usw. durchgeführten Stichprobe und dem daraus errechneten Durchschnittswert auf ein universal gültiges Gesetz “Höhe der Pflanze Habichtskraut” zu schließen.⁴⁴⁸ Man könnte zuerst annehmen, daß diese Zufallsstichprobe für 280 Pflanzen, die sich auf der Wiese befinden, repräsentativ ist. Aber: “Eine Wissenschaft von 280 Pflanzen auf einer hic et nunc vorhandenen Wiese gibt es nicht.”⁴⁴⁹

⁴⁴⁶ So existierte z. B. keine Theorie, die festlegte, daß Monate und nicht Quartale oder Halbjahre einen Einfluß auf die Geburtenrate ausübten.

⁴⁴⁷ Das Beispiel des Habichtskrauts nach Menges (1959a).

⁴⁴⁸ Auch wenn von einer hypothetisch unendlichen Grund-/Zielgesamtheit ausgegangen wird, ist die Population, aus der eine Stichprobe gezogen wird, stets eine endliche, begrenzt in Zeit und Raum. Schlüsse, die aufgrund einer Stichprobe gezogen werden, sind dann nichts anderes als eine Deskription der entsprechenden Kenngrößen der Grund-/Zielgesamtheit. Vgl. Rutsch (1985, S. 196). Auf diesen Aspekt hat insbesondere die – hier nicht behandelte – “kontinentale Schule der mathematischen Statistik” hingewiesen. Ihr letzter Vertreter, O. Anderson, definierte den Begriff der (sozialstatistischen) Wahrscheinlichkeit daher auch konsequent als Anteil in einer übergeordneten Grundgesamtheit.

⁴⁴⁹ Menges (1959a, S. 15).

Die Frage ist also letztendlich, was mit "allgemeinen Bedingungen" gemeint ist. Dies können nur andere theoretische Größen wie Lichtbedingungen, Klimaverhältnisse, Bodenbeschaffenheit usw. sein, die wir in der vorliegenden Situation messen, beeinflussen und damit kontrollieren können. In dieser Situation liegt nun auch das Spezifische der naturwissenschaftlichen Statistik: "Einen Prüfstein für die Richtigkeit des Wahrscheinlichkeitsschlusses haben wir allemal im Experiment. Denn wenn wir unter den gleichen allgemeinen Bedingungen, wie sie beim Zustandekommen der für den [...] Schluß bestimmten Masse geherrscht haben, Experimente durchführen, so müssen die möglichen realen Folgen mit den ihnen entsprechenden Wahrscheinlichkeiten wieder zum Vorschein kommen, und dies, so oft wir nur die Versuche wiederholen [...]"⁴⁵⁰

Wir sehen somit bereits zu diesem Zeitpunkt eine Ausrichtung, in der der Möglichkeit des Experiments eine fundamentale Rolle zukommt. Ungeachtet dessen schufen – an Galton anknüpfend – F. Y. Edgeworth und vor allem K. Pearson eine Wissenschaft der Statistik, die sich als Fach verselbständigte. Seither existiert die Vorstellung, daß es einen Methodenkanon "der" Statistik gibt, der prinzipiell variier- und erweiterbar, aber weitgehend unabhängig von dem Analyseobjekt ist. Der Korrelationskoeffizient etwa als zusammenfassende Größe für metrische Daten erschien universell als sinnvolle Maßzahl für den Zusammenhang zweier oder mehrerer Größen.

Die eigentliche Öffnung statistischer Methoden für die Sozialwissenschaften erfolgte, gegen den Widerstand von K. Pearson, durch G. U. Yule. Diese Übernahme war bereits durch eine Reihe von pragmatischen Erwägungen charakterisiert (Art der funktionalen Beziehung, Minimierungskriterium), die in der Folgezeit zumeist als selbstverständlich hingenommen wurden. Yule erweiterte Galtons Korrelations- und Regressionsansatz in eine multivariate Richtung, übernahm jedoch nicht dessen theoretische Rechtfertigung für ein solches Vorgehen. Rein formal handelte es sich bei diesem multiplen Korrelations- und Regressionsansatz um die gleiche Situation wie in der Astronomie und Geodäsie: eine funktionale Beziehung und mehr Beobachtungen als Unbekannte. Yule übernahm daher das von Laplace und Gauss vollständig entwickelte Instrumentarium. War eine solche Vorgehensweise jedoch zu rechtfertigen? In der Astronomie und Geodäsie bestand das Problem z. B. darin, eine Länge aus einer Kombination von Längen und Winkeln zu ermitteln. Wenn die Größen fehlerfrei gemessen wurden, bestand an der Richtigkeit des "Modells", des funktionalen Zusammenhangs, kein Zweifel, denn die Gleichung war eine mathematische (oder geometrische) Identität.⁴⁵¹ Doch konnte man dies auch von der Beziehung zwischen

⁴⁵⁰ Ebda., S. 19.

⁴⁵¹ Die funktionalen Beziehungen, die in diesem Zusammenhang untersucht wurden, sind recht komplizierter Natur, so daß wir hier nicht näher darauf eingehen wollen. Die Problemstellung ist jedoch vergleichbar mit folgendem vereinfachten Beispiel: Der Satz des Pythagoras besagt, daß in einem rechtwinkligen Dreieck die Summe der Quadrate über den Katheten a und b gleich dem Quadrat der Hypothenuse c ist. Angenommen, man ist für ein physikalisch existierendes Objekt an der Größe c interessiert, kann aber nur a und b messen. Dann kann man, auch wenn a und b meßfehlerbehaftet sind, mittels des Kleinstquadratverfahrens oder

- Erbsengrößen und Lichteinfluß (Galton),
- der Körpergröße von Kindern und Eltern (Galton) oder gar
- den Einfluß von “Determinanten” auf den Pauperismus (Yule)

behaupten? Keine biologischen, geschweige denn sozioökonomische Theorien konnten und können Aussagen dieser Art treffen.

Unseres Erachtens läßt sich nicht entscheiden, ob Yule konzeptuell von Meßfehlern (analog zur Physik) oder von Verteilungen der Variablen selbst (analog zur Biologie) ausging.⁴⁵² Eine Klärung dieser Frage wird dadurch erschwert, daß zu diesem Zeitpunkt bereits ein stillschweigender Übergang von einer inversen zu einer häufigkeitstheoretischen Sicht vollzogen worden war. Laplace, Gauss, Galton und Edgeworth hatten bezüglich inferenzstatistischer Aussagen “bayesianisch” argumentiert. Für Gauss etwa war der wahrscheinlichste Parameterwert das Maximum der Likelihoodfunktion, da er wie Laplace vom Prinzip des unzureichenden Grundes und somit von einer A-priori-Gleichverteilung ausging. Pearson folgte dagegen – meistens – einer Stichprobenkonzeption, und auch Yule bewegte sich in diesem Rahmen, wenngleich er der Frage der Inferenz generell wenig Bedeutung beimaß.

Die weitere Entwicklung der statistischen Methodologie wurde spätestens seit K. Pearson und R. A. Fisher von der Biologie ausgehend durch die Anwendungsmöglichkeit in den Naturwissenschaften geprägt. Die Probleme des Inferenzschlusses in der Biologie durch die Abhängigkeit von den während der Stichprobenziehung waltenden Umständen versuchte Fisher durch das *Design of Experiments* zu lösen. Von der Biologie ausgehend wurde die statistische Methodologie auf nahezu alle anderen empirischen Wissenschaften übertragen; oftmals mit zweifelhafter epistemologischer Fundierung:

“Richtig gepaßt hat die Fisher’sche Methodologie aber stets nur auf die klassische Biologie und eng benachbarte Disziplinen wie die Agronomie. Schon in der Psychologie oder der Medizin, erst recht in den Wirtschafts- und Sozialwissenschaften, aber gab und gibt die Fisher’sche Methodologie vielfach Antworten auf nicht gestellte Fragen, während die eigentlichen Fragen unbeantwortet bleiben.”⁴⁵³

Fisher entwickelte statistische Methoden im Hinblick auf die Möglichkeit, den Umfang der Daten in einem Experiment beliebig vermehren, vor allem aber beliebig oft Experimente mit geringer Fallzahl durchführen zu können.⁴⁵⁴ Nur dann

der Maximum-Likelihood-Methode unter der Annahme normalverteilter Fehler $c^2 = (a^2 + b^2)$ ermitteln.

⁴⁵² Pearson als Biometriker wollte lediglich die letzte Analogie gelten lassen. Nur wenn die Daten eine gemeinsame Normalverteilung aufwiesen, war man berechtigt, die Relationen durch einzelne Zahlen (Korrelations- oder Regressionskoeffizienten) auszudrücken, denn in solchen Fällen konnte die gesamte Verteilung rekonstruiert werden. Sonst müsse man die vollständige gemeinsame Häufigkeitsverteilung, die “frequency surfaces”, betrachten.

⁴⁵³ Menges (1980b, S. 182f).

⁴⁵⁴ Lancaster (1988, S. 708) sieht in dieser speziellen Anwendungsbezogenheit auf der anderen Seite gerade den Erfolg Fishers: “His field of application was fortunately chosen; the results

ergaben Kriterien wie Suffizienz, Erwartungstreue, Effizienz und Konsistenz einen Sinn. Die Forderung, daß eine "Statistik" (eine Schätzung) suffizient sein, also die gesamte in den Daten zur Verfügung stehende Information ausnutzen sollte, erweist sich zweifellos als sinnvolle Forderung, auch und gerade wenn man nicht in der Lage ist, Experimente durchzuführen. Stegmüller fragt jedoch zu Recht:

"Wie aber steht es mit den drei übrigen Merkmalen der Erwartungstreue, Effizienz und Konsistenz? Könnte man zwingende Gründe zugunsten dieser Eigenschaften vorbringen, so wären dies vermutlich indirekt auch Gründe dafür, die Güte im absoluten Sinn der Güte im Stützungssinn vorzuziehen [...]. Nun scheint es aber, daß überzeugende Gründe niemals vorgebracht worden sind, sondern statt dessen nur mehr oder weniger vage Appelle an die Intuition vorgenommen worden sind. Es liegt daher nahe, den Spieß umzudrehen und umgekehrt zu fragen, ob sich Bedenken gegen die Auszeichnung von Schätzfunktionen mit diesem Merkmal vorbringen lassen, und wenn ja, welcher Art diese Bedenken sind. Zunächst muß hervorgehoben werden, daß überhaupt nicht einzusehen ist, warum ein Merkmal der Optimalität auf lange Sicht [d. h. für wiederholte Stichprobenziehungen, T. R.] auch ein vorteilhaftes Merkmal für einen konkreten Einzelfall sein soll. Das gilt um so mehr, als die long-run-Betrachtung nicht einmal eine lange Sicht im menschlichen Sinn, sondern nur im Sinn mathematischer Konvergenz darstellt."⁴⁵⁵

Die weitere Entwicklung, insbesondere die Theorie von Neyman/Pearson und Wald, hat diese Diskrepanz eher noch vergrößert. Nun war statistische Inferenz auf die Aufstellung von Richtlinien für ein Verhalten auf lange Sicht reduziert.⁴⁵⁶ Durch die Neyman-Pearson-Theorie wurde keine Klärung strittiger epistemologischer Fragen erreicht, sondern nur eine eindeutige Stellungnahme bezogen. Vielleicht erklärt sich ihr Erfolg auch damit, daß andere Positionen (K. Pearson, Fisher) solch eine Eindeutigkeit vermissen ließen.

Seither dominierte jedenfalls eine angeblich objektive, häufigkeitstheoretische Inferenzsicht, wobei sich immer wieder Elemente fanden, die implizit bayesianischer Natur waren. Neobayesianische Standpunkte, die sich etwa zur gleichen Zeit entwickelten, konnten sich dennoch nicht durchsetzen. Es ist jedoch bemerkenswert, daß die moderne subjektivistische *Wahrscheinlichkeitstheorie* nicht von Sozialwissenschaftlern begründet wurde, die einzelne Voraussetzungen oder Implikationen dieser auf lange Sicht angelegten experimentellen Inferenz als problematisch ansahen, sondern ausnahmslos von Mathematikern (Ramsey, de Finetti, Savage) oder Geophysikern (Jeffreys), die logische Probleme in den vorherrschenden häufigkeitstheoretischen Ansätzen erblickten.

had immediate application with assessable economic value; realistic simplifying hypotheses, such as the normality of error and their mutual independence, could be made; there were few entrenched doctrines to combat; the price of experiment was low; there were no ethical problems." Insofern ist die moderne Statistik in weiten Bereichen genauso kontextbezogen wie die Entwicklung der Fehlertheorie.

⁴⁵⁵ Stegmüller (1973, S. 205f). In diesem Zusammenhang drängt sich auch das berühmte Zitat von Keynes (1924, S. 121) auf: "In the long run we are all dead."

⁴⁵⁶ Bemerkenswert ist in diesem Zusammenhang, daß sich die "small sample theory" auf eine häufigkeitstheoretische Definition beruft, deren Fundierung vor allem auf von Mises zurückgeht, dieser jedoch eine solche Theorie ablehnte.

Für diese entstehende neobayesianische Inferenz war eine alternative Definition des Wahrscheinlichkeitsbegriffes entscheidend. Um den Gegensatz besser zu verdeutlichen, wollen wir uns zunächst dem klassischen Wahrscheinlichkeitsbegriff und seinen Schwächen zuwenden.

Stegmüller zählt acht in der Literatur vorgebrachte Einwände gegen die Häufigkeitstheoretische, auf von Mises zurückgehende Definition auf und hält zumindest den letzten für "tödlich": Er verwechsle praktische Sicherheit mit logischer Notwendigkeit.⁴⁵⁷ Eine besondere Schwäche dieses Wahrscheinlichkeitsbegriffes wurde in der Ablehnung von Einzelwahrscheinlichkeiten gesehen. Nach von Mises Definition war es unmöglich, etwa die Wahrscheinlichkeit für eine bestimmte Augenzahl für den Wurf mit einem konkreten Würfel an einem konkreten Ort anzugeben. Noch deutlicher wird die Problematik einer solchen Definition an einem anderen Beispiel: Wenn der Anteil der männlichen Personen,⁴⁵⁸ die an Krebs erkranken, bei $x\%$ liegt, der Anteil bei rauchenden Personen bei $y\%$, dann stellt sich die Frage, wie groß demnach die Wahrscheinlichkeit für einen männlichen Raucher ist, an Krebs zu erkranken. Die Gefahr einer Krebserkrankung hängt mit Sicherheit von einer großen Zahl von Faktoren ab, dennoch wird die individuelle Wahrscheinlichkeit nach dieser Logik nur dann als Kombination der relativen Häufigkeiten der an Krebs bereits erkrankten rauchenden und männlichen Personen aufgefaßt werden können, wenn man alle anderen Faktoren als konstant betrachtet. Dies bedeutet nichts anderes, als einer individuellen Person aufgrund 1) unserer Kenntnis von relativen Häufigkeiten und 2) unserer Unkenntnis aller Faktoren, die für die konkrete Person relevant sind, eine individuelle Wahrscheinlichkeit zuzuschreiben.⁴⁵⁹

⁴⁵⁷ Vgl. Stegmüller (1973, S. 32ff, insbesondere S. 37). In der Tat ist die Konvergenz der relativen Häufigkeit gegen einen als statistische Wahrscheinlichkeit bezeichneten Wert p nicht logisch notwendig, sondern nur "P-fast sicher". Das bedeutet, daß die – statistische – Wahrscheinlichkeit dieser Konvergenz gleich 1 ist.

⁴⁵⁸ Wir lassen hier vorerst die Problematik der Grundgesamtheit außer acht.

⁴⁵⁹ Der Begriff der Zufälligkeit läßt sich daher im ontologischen Sinne wohl nur in sehr begrenzten Bereichen wie etwa der Quantenmechanik aufrecht erhalten. Dort kann man eine Wahrscheinlichkeitsverteilung prinzipiell nur beobachten, nicht erklären. Alle weiteren Ereignisse bauen darauf auf. Die Bezeichnung von Ereignissen als "zufällig" ist lediglich Ausdruck unserer Unkenntnis und nicht der Abwesenheit physikalischer Vorgänge, die diese Realisationen determiniert haben. Stegmüller (1973, S. 290) liefert ein anschauliches Beispiel. Für das Ereignis "Geburt eines Knaben" läßt sich a priori eine bestimmte Wahrscheinlichkeit angeben. Dennoch ist dieses Ereignis im ontologischen Sinne nicht zufällig. Bei Kenntnis des Typs der Samenzelle ist auch das Geschlecht des Kindes (vor dem Ereignis "Geburt") bekannt: "Die statistische Analyse bildet [...] im vorliegenden Fall nur eine *Oberflächenanalyse*." Hervorhebung im Original. In diesem Sinne läßt sich prinzipiell – mit Ausnahme des quantenphysikalischen Bereiches – jedes Ereignis kausal (deterministisch) erklären. Auch ein Würfelwurf ist prinzipiell prognostizierbar, wenn nur Rahmen- und Randbedingungen bekannt sind. In der Regel werden aber zwar die Rahmenbedingungen (d. h. die "Versuchsanordnung" oder die "Oberhypothesen") identisch sein, die Randbedingungen jedoch nicht. Für die Situationen, in denen statistische Gesetze empirisch untersucht werden, sind lediglich die *Häufigkeitsverteilungen dieser Randbedingungen* feststellbar. Popper (1994, S. 159) spricht daher auch zu Recht von zufallsartig anstelle von zufällig.

Aus dieser Problematik heraus hat etwa K. R. Popper, einer der vehementesten Gegner des Subjektivismus, seine Auffassung des Wahrscheinlichkeitsbegriffes – allerdings vor allem auf die Probleme der Physik bezogen – im Laufe der Jahre zu einer “Propensity”-Theorie hin entwickelt, der sich im wesentlichen auch Stegmüller anschließt.⁴⁶⁰ Stegmüller sieht sich gezwungen, Wahrscheinlichkeit als einen theoretischen Begriff, “eine nicht explizit definierbare *theoretische Disposition* physikalischer Systeme” zu bezeichnen⁴⁶¹ und gelangt zu der Einsicht, daß der Versuch zwecklos sei, “den Begriff der statistischen Wahrscheinlichkeit mittels bereits verfügbarer logischer und empirischer Begriffe zu definieren.” Damit erweise sich auch der Streit zwischen Objektivisten und Subjektivisten, die beide stillschweigend von dieser Oberhypothese ausgingen, eigentlich als Pseudogegensatz:

“Wenn man an einem objektiven Begriff der statistischen Wahrscheinlichkeit festhalten möchte, dann kann seine Einführung nicht durch die von der Limestheorie vorgeschlagene definitorische Zurückführung auf den Begriff der relativen Häufigkeit erfolgen, es sei denn bei Strafe der Kapitulation vor dem Subjektivismus. Es besteht noch immer die Möglichkeit eines ganz andersartigen Ausweges, der keine solche Kapitulation impliziert. Er besteht in der Preisgabe der reduktionistischen Grundannahme, welche besagt, der Begriff der statistischen Wahrscheinlichkeit müsse mittels bereits verfügbarer mathematischer und empirischer Begriffe *explizit definierbar* sein.”⁴⁶²

Es ist unserer Ansicht nach allerdings fragwürdig, ob der Rückzug auf eine solche Anti-Definition einer bayesianischen, subjektivistischen Interpretation überlegen ist. Andererseits muß jedoch festgestellt werden, daß bis in die Gegenwart hinein keine Einigung erzielt werden konnte und auch für die nächste Zukunft eine solche Klärung wohl nicht zu erwarten ist, wie dies noch vor kurzem C. Howson feststellte:

“It would be foolhardy to predict that philosophical probability has entered a final stable phase; surveys of the field tend to have useful lifetimes of a decade or so, at most two. It would also probably be incorrect to pretend that there is likely in the near future to be any settled consensus as to which interpretations of probability make viable and useful theories, and which are dead ends.”⁴⁶³

⁴⁶⁰ Vgl. Popper (1994, Kap. 8, 10 (insbesondere S. 219, Anm. *1), Anhang *II - *VII, *IX, *XIII, *XVII), Popper (1995), Popper (1996) sowie Stegmüller (1973, S. 245ff). Zum gegenwärtigen Stand der wahrscheinlichkeitstheoretischen Diskussion in der Physik siehe Jaynes (1996).

⁴⁶¹ Stegmüller (1973, S. 58). Hervorhebung im Original. Er bemerkt (ebda.), daß R. B. Braithwaite bereits bei R. A. Fisher eine solche Konzeption vermutet habe. Derartige Vorstellungen lassen sich aber bereits erheblich früher festmachen. So findet man bei de Moivre die folgende Definition: “Chance, as we understand it, supposes the Existence of things, and their general known Properties: that a number of Dice, for instance, being thrown, each of them shall settle upon one or other of its Bases. After which, the Probability of an assigned chance, *that is of some particular disposition of the Dice*, becomes as proper a subject of Investigation as any other quantity or Ratio can be.” Aus: A. De Moivre, *The doctrine of chances*, 3. Aufl., London 1756, zitiert nach Schneider (1988, S. 133). Hervorhebung von uns.

⁴⁶² Stegmüller (1973, S. 39). Hervorhebungen im Original.

Bayesianische *Inferenzkonzeptionen* sind jedoch nicht auf einen subjektiven, die A-priori-Wahrscheinlichkeit formalisierenden Teil beschränkt, sondern verbinden diesen mittels des Bayes-Theorems mit der "Evidenz der Daten", die in der Likelihoodfunktion formalisiert wird. Die Likelihoodfunktion spielte bereits bei Bernoulli, Laplace und Gauss eine wichtige Rolle. Ihre Bedeutung als zentrales Element statistischer Inferenz wurde vor allem von A. Birnbaum betont, der in diesem Zusammenhang den Begriff des Likelihood-*Prinzips* einführte. Der Hauptunterschied zwischen dem Likelihood-Prinzip und dem Häufigkeitsprinzip läßt sich als Frage formulieren: Kann man Evidenz über einen Parameter anhand eines spezifischen Datums (einer "Stichprobe") erhalten? Anhänger der Häufigkeitskonzeption (v. a. J. Neyman) betonen, daß man die Leistung einer Prozedur nur beurteilen könne, wenn man sie wiederholt durchführe und langfristige Durchschnitte betrachte.

Falls man jedoch nicht in der Lage ist, Experimente durchzuführen, sondern Schlüsse anhand von vorliegenden, nicht kontrolliert wiederholbaren Daten gezogen werden sollen, so ist die Relevanz einer solchen Konzeption ernsthaft zu hinterfragen. Falls die Wiederholbarkeit rein hypothetisch ist, sollte dies auch explizit als (subjektive) Überzeugung und nicht als objektive Möglichkeit definiert werden. Für solche Situationen erscheint uns daher eine Definition der Wahrscheinlichkeit als Grad der Überzeugung, die einem Parameterwert zugesprochen wird, angemessener. Die Gewichtung und Revision dieser Überzeugung mit der Evidenz der vorliegenden, nicht hypothetischen Daten, die über die Likelihoodfunktion vorgenommen wird, ist darüber hinaus unserer Ansicht nach logisch konsistent, zumal sie nicht von asymptotischen Verallgemeinerungen abhängig ist. Wir wollen uns hier dem Urteil D. Lindleys anschließen:

"The present position in statistical inference is historically interesting. The bulk of practitioners use well-established methods like least squares, analysis of variance, maximum likelihood and significance tests: all broadly within the Fisherian school and chosen for their proven usefulness rather than their logical coherence. If asked about their rigorous justification most of these people would refer to ideas of the NPW [Neyman-Pearson-Wald, T. R.] type; least-square estimates are best, linear unbiased; F-tests have high power and maximum likelihood values are asymptotically optimal. Yet these justifications are far from satisfactory: the only logically coherent system is the Bayesian one which disagrees with the NPW notions, largely because of their violation of the likelihood principle."⁴⁶³

Kommen wir nun zum nächsten Problemkreis, dem Aspekt der Modellkonstruktion. Diesem Aspekt ist besonders Rechnung zu tragen, wenn das Modell weder durch eine Theorie vollständig vorgegeben noch durch Experimente evaluiert werden kann. Ein "klassischer" Inferenzschluß kann sich prinzipiell nur auf eine (endliche oder unendliche) Grundgesamtheit beziehen, aus der alle Elemente mit gleicher Wahrscheinlichkeit gezogen werden können. Die Gesamtheit, die das Ziel einer Untersuchung ist, muß nicht notwendigerweise identisch sein mit der Gesamtheit, aus der eine Stichprobe gezogen wurde. Zahlreiche und hinlänglich

⁴⁶³ Howson (1995, S. 27).

⁴⁶⁴ Lindley (1991, S. 493).

bekannte Gründe – selection bias usw. – können dies verhindern. Dieses Problem ist um so gravierender, je weniger Einflußmöglichkeiten man auf die Auswahl der Elemente hat, also insbesondere im Falle *modellbasierter Inferenz*.⁴⁶⁵

Für unseren Zusammenhang sind insbesondere Modelle im Rahmen der Zeitreihenanalyse von Bedeutung. Wir hatten festgestellt, daß zunächst, wiederum in Anlehnung an die Astronomie, ein deterministisches Komponentenmodell auch auf sozioökonomische Phänomene übertragen wurde. In diesem Rahmen fand z. B. die Korrelation als Maßzahl für den Zusammenhang zweier Entwicklungen bereits selbstverständliche Verwendung, wobei insbesondere G. U. Yule auf Probleme hinwies.

Die Theorie stochastischer Prozesse entwickelte sich dagegen in einem komplexen Konglomerat unterschiedlichster Wissenschaften. Für den Anwendungsbereich wurden die Arbeiten von Yule und Slutsky grundlegend. Diese Modelle waren noch durch umfangreiche Überlegungen über Kausalitäten und Ursachen motiviert. Im weiteren Verlauf überwog jedoch bei der Behandlung stochastischer Prozesse die Lösung formaler Probleme. Schrittweise wurden sehr spezielle Modellerweiterungen, Vorstellungen und Methoden entwickelt; nicht, weil sie sich logisch aus dem Problem heraus begründeten, sondern weil man schon bei einer minimalen Erweiterung des Anwendungsbereiches auf technische Probleme stieß, deren Lösung Energien absorbierte. Nicht ohne Bedeutung dürfte darüber hinaus eine in solchen Situationen oftmals entstehende Eigendynamik sein: Man entfernt sich zunehmend vom eigentlichen Problem und entwickelt einen Ehrgeiz, den *formalen* Teil möglichst allgemein und logisch stringent zu bewältigen.

Als Box und Jenkins an die Tradition von Yule und Slutsky anknüpften, war von den ursprünglichen Intentionen nicht mehr viel übrig. Ausgehend von Optimierungsproblemen für technische Geräte (Katalysatoren) suchten sie nach effizienten Konstruktionsmechanismen, die für vorhandene Daten möglichst gute Anpassungen und Prognosen liefern sollten, gleichgültig, *weshalb* sich z. B. in einer konkreten Situation ein MA(1) und kein AR(1)-Modell als angemessen erwiesen hatte. Betrachtet man die Entwicklung der Zeitreihenanalyse vor dem Hintergrund der Entstehung der Inferenzstatistik, so ist es erstaunlich, daß *evolutorische Aspekte* überhaupt keine Rolle spielten. In der Analyse von Zeitreihen wurde davon ausgegangen, daß zwar die *Beobachtungen* zeitabhängig sein konnten, nicht jedoch die *Mechanismen*. Lautete aber nicht eine immer wieder vorgebrachte Kritik, insbesondere in bezug auf gesellschaftliche Anwendungen, daß Relationen und Zusammenhänge selbst einem zeitlichen Wandel unterworfen waren? Erst die Arbeit von Brown, Durbin und Evans (1975) schaffte auf breiterer Ebene ein Bewußtsein dafür, daß die Parameterkonstanz zumindest überprüft werden sollte, auch wenn die Autoren keine zeitabhängige Parametermodellierung als Alternative formulierten.⁴⁶⁶

⁴⁶⁵ Nach der Definition von Koch/Gillings (1983), siehe oben, S. 105f.

⁴⁶⁶ Wir wollen nicht bestreiten, daß die Problematik der Parameterkonstanz bereits früher vereinzelt diskutiert wurde. Für die Entwicklung der statistischen Methodologie insgesamt hat dies jedoch keine Rolle gespielt.

Zwei Aspekte sind hier hervorzuheben: Erstens hat sich die statistische Zeitreihenanalyse bis heute von dieser Problematik weitgehend unbeeinflusst gezeigt. Nach wie vor dominieren hier die von Box und Jenkins eingeführten ARIMA-Modelle, in denen die Struktur, der "gesetzmäßige" Teil, zeitlos ist. Zweitens hat seit den sechziger Jahren in der ingenieurwissenschaftlichen Literatur als explizite Modellierung zeitabhängiger Parameter der sog. Kalmanfilter starke Verbreitung gefunden. Bis heute hat jedoch keine systematische Einordnung dieses Verfahrens in statistische Methodologien stattgefunden; Anwendungen stehen lediglich vereinzelt da.⁴⁶⁷ Nach überwiegender Einschätzung, der wir uns anschließen wollen, drängt sich hier eine bayesianische, sequentielle Argumentation geradezu auf. Die zu einem bestimmten Zeitpunkt t vorhandene A-priori-Information (d. h. die Information bis zum Zeitpunkt $t-1$) wird durch die neu hinzukommende Information, konkret durch ihre Likelihoodfunktion, mittels des Bayes-Theorems zur A-posteriori-Information zum Zeitpunkt t kombiniert. Konzeptionell ist somit eine viel angemessenere Inferenz möglich, falls der zu analysierende Prozeß nicht als zeitlos mechanisch aufgefaßt werden kann. Eine solche Situation dürfte außerhalb der Naturwissenschaften eher die Regel als die Ausnahme darstellen. Schließlich muß jedoch betont werden, daß auch die Vorstellung variabler Parameter eine oftmals notwendige, letzten Endes aber wohl selten hinreichende Modellerweiterung darstellt, denn auch die Modellkonstanz für die Entwicklung der Parameter kann wiederum hinterfragt werden. Es bleibt die Frage, auf welcher Ebene die Gesetzmäßigkeit postuliert werden muß:

"Wissenschaftliche Gesetze formulieren stets Aussagen über 'zeitinvariante Strukturen' in einer scheinbar ungeordneten Welt, und die Feststellung, daß die Koeffizienten eines (stochastischen) Modells für eine Gruppe [...] Variabler einer Veränderung unterliegen, führt zwangsläufig auf die Frage, worauf diese Veränderungen zurückzuführen sind. Man wird versuchen, ein Bewegungsgesetz zu finden. Die Konstanz der Parameter wird also nur auf eine höhere Ebene verschoben."⁴⁶⁸

Ein weiterer Aspekt, dem wir uns zugewandt hatten, betraf die Explorative Datenanalyse. Derartige Analysekonzeptionen versuchen, dem Modellkonstruktionsprozeß explizit Rechnung zu tragen. Sowohl J. W. Tukey als auch G. E. P. Box betonten den *iterativen Charakter* der Datenanalyse. Ein solches Vorgehen ist jedoch in einer nicht-experimentellen modellbasierten Inferenz problematisch.⁴⁶⁹ Wenn Tukey forderte, die vergebliche Hoffnung aufzugeben, daß die Datenanalyse in einem deduktiven System begründet werden könne, dann bleibt als Alternative unserer Ansicht nach entweder die Beschränkung auf eine reine Deskription oder

⁴⁶⁷ Es ist außerdem darauf hinzuweisen, daß in nicht-ingenieurwissenschaftlichen Problemstellungen oftmals zusätzliche Probleme vorhanden sind. Größen, die in jenem Kontext theoretisch vorgegeben sind, müssen hier geschätzt werden.

⁴⁶⁸ Schneider (1986, S. 42), Hervorhebung im Original.

⁴⁶⁹ Man sollte sich darüber hinaus fragen, ob ein solches Vorgehen tatsächlich iterativ oder strenggenommen nicht eher *pfadabhängig* ist. Die einzelnen Schritte hängen auch vom Ereignis vorheriger Schritte ab; bei einem Rückschritt sind aber nicht mehr alle ehemaligen Alternativen offen. Siehe dazu auch Rutsch (1985, S. 204).

die Akzeptanz bayesianischer Inferenz. Wenn man sich auf die klassische Häufigkeitstheoretische Interpretation bezieht, so muß das Modell vor der Analyse der Daten feststehen, ansonsten ergeben Signifikanztests selbst im Rahmen der klassischen Inferenzlogik keinen Sinn und können allenfalls instrumentalistisch verstanden werden. Wenn diese Tests aber nicht die innere Bedeutung haben, die ihnen von den Verfechtern der klassischen Inferenzkonzeption zugesprochen wird, warum sollte man dann nicht von vornherein ihren subjektiven Charakter (denn sie sind von subjektiven Entscheidungen im Modellbildungsprozeß abhängig) akzeptieren? Dies gilt um so mehr, als für Box ohnehin alle Modelle falsch sind. Die logische Konsequenz muß lauten, daß damit alle Signifikanztests falsch sind. In diesem Sinne ist auch das in hohem Maße explorative Vorgehen in der Zeitreihenanalyse nach der "Box-Jenkins-Methodologie" zu verstehen. Überraschenderweise spielten in diesem Zusammenhang bayesianische Interpretationen nur eine untergeordnete Rolle, obwohl Box im Rahmen seiner allgemeinen Datenanalysekonzeption mehrfach die Forderung nach einer bayesianischen Denkweise gestellt hat.

Erst in jüngster Zeit hat sich die theoretische Statistik intensiver mit der Problematik der Auswirkung von iterativen Vorgehensweisen und Modellselektionsprozeduren befaßt und ansatzweise begonnen, Lösungen zu entwickeln. Mittlerweile hat man auch hier realisiert, daß die Frage nach der besten Schätzung unter der Bedingung, daß das Modell "richtig" ist, in vielen Situationen von erheblich geringerer Bedeutung ist als eine Abschätzung der Modellunsicherheit.

Draper vergleicht die Situation der Nichtberücksichtigung von Modellunsicherheit mit der Schätzung von z -Werten bei unbekannter Varianz vor der Einführung der t -Verteilung durch W. Gosset. Man benutzte die vorliegenden Daten zur Schätzung der Varianz der Grundgesamtheit und setzte diese in die Formel anstelle der nicht bekannten Varianz der Grundgesamtheit ein, berücksichtigte dies jedoch nicht als Schätzung. Auch nach der Einführung der t -Verteilung durch Gosset mußten zehn bis zwanzig Jahre vergehen, bevor ihre Anwendung allgemeine Verbreitung fand. Es mag dahingestellt sein, ob das Problem der Berücksichtigung von Modellunsicherheit in der Tat durch die vorhandenen Ansätze bereits gelöst ist; auf jeden Fall sollte diesem Aspekt zukünftig – und nicht erst in zehn oder zwanzig Jahren – stärkere Beachtung geschenkt werden. Soweit wir sehen, existieren zur Zeit vor allem innerhalb bayesianischer Ansätze geeignete Lösungsansätze. Dies scheint uns auch der einzig mögliche Weg zu sein.

“Unfortunately, all the hard scientific evidence since Newton has shown that there are no universal laws for random phenomena (the law of large numbers and other mathematical artifacts notwithstanding) because the observed randomness is fundamentally influenced by the *system* through which the uncertainty becomes accessible for observations. [...] The aspiration of Haavelmo to give a solid foundation to econometrics by dogmatic application of probability theory has not been fulfilled [...], no doubt because probability theory has nothing to say about the underlying system-theoretic problems.”

Rudolf Kalman (1982b, S. 172, 194)

B. Konzepte der Ökonometrie

1. Allgemeines

Das Aufgaben- bzw. Tätigkeitsgebiet der Ökonometrie ist hauptsächlich durch die folgenden, über eine reine Beschreibung empirischer Phänomene hinausgehenden Bereiche definiert worden: 1. die Überprüfung von Hypothesen, 2. die Schätzung von Parameterwerten eines Modells anhand von Daten und 3. die Prognose zukünftiger oder nicht realisierter Zustände.¹

Mit der Durchführung dieser Aufgaben ist die Ökonometrie in den letzten sechzig Jahren zu einer tragenden Säule der Wirtschaftswissenschaften geworden und hat hier im Laufe der Entwicklung wie kaum ein anderes Gebiet an Bedeutung gewonnen. Dies zeigt sich nicht zuletzt daran, daß 1969 der erste Nobelpreis für Wirtschaftswissenschaften an zwei Gründungsväter der Ökonometrie, R. Frisch und J. Tinbergen, verliehen wurde und ihnen 1975 (T. Koopmans), 1980 (L. R. Klein) und 1989 (T. Haavelmo) weitere Ökonometriker folgten.

Nun wäre die Durchführung der genannten drei Tätigkeiten, die auf den ersten Blick ohne weiteres mit den Mitteln der theoretischen Statistik bewältigt werden könnte, alleine wohl kaum eine nobelpreiswürdige Leistung, und es ist daher kaum verwunderlich, daß noch weitere Bereiche den Kernbestandteilen der Ökonometrie zuzurechnen sind: zumindest die Entwicklung spezifischer Methoden gehört dazu.² Zwar wurden aus dem Bereich der theoretischen Statistik methodische Entwicklungen (und vor allem epistemologische Grundlagen) übernommen, jedoch entwickelte die Ökonometrie auch eigenständig Verfahren und Konzepte. Denn während die Statistik im Rahmen und für experimentelle Wissenschaften entwickelt wurde, muß sich die Ökonometrie überwiegend auf das passive Beobachten von Daten beschränken, die zudem (in der Regel) nur in sehr geringem Umfang vorhanden sind.³ Auch der Konstruktion von Modellen kommt naturgemäß in der Ökonometrie eine größere Rolle zu als in vielen anderen Wissenschaft-

¹ Vgl. Menges (1959b, S. 615). Ähnliche Definitionen finden sich in den einschlägigen Lexikonartikeln.

² Vgl. Brinkmann (1970, S. 205). Er fügt dem noch die Operationalisierung sowie die Gewinnung von Daten hinzu.

³ Leserer (1986, S. 99) spricht in diesem Zusammenhang vom “Elend ökonometrischen Schätzens”.

ten, denn die Ökonomie bietet oftmals eine ganze Reihe von Theorien, die anhand desselben, in der Regel unveränderbaren Datenmaterials getestet und gegebenenfalls modifiziert werden müssen.

Diese eigenständige Entwicklung hat in den letzten Jahrzehnten zu einem Methodenarsenal geführt, das auf monographische Weise kaum noch überblickt werden kann⁴ und selbst Tinbergen offen zugeben ließ, daß er einen Großteil der Beiträge der führenden Zeitschrift *Econometrica* aus Zeitgründen nicht mehr lesen könne.⁵ Eine Tendenz läßt sich dennoch feststellen: Es ist eindeutig, daß die ursprüngliche Konzeption der Theorienüberprüfung zumindest um das Konzept der Datenanalyse ergänzt, wenn nicht gar dadurch ersetzt worden ist.⁶

Zu den Gründen für diese Entwicklung. Auch hier erscheint uns zunächst eine historische Betrachtung maßgeblicher Konzepte hilfreich. Mit diesem Ansatz befinden wir uns in guter Gesellschaft, denn auch die Ökonometrie hat mittlerweile ein "Geschichtsbewußtsein" entwickelt, und es mehren sich Ansätze, sich methodologischen Aspekten durch eine wissenschaftshistorische Betrachtungsweise zu nähern, wobei die extensive Spezialisierung des Faches und eine dadurch verursachte Orientierungssuche auch eine Rolle spielen dürften. So fragte etwa T. Tewes vor wenigen Jahren:

"Hat die Ökonometrie mehr als ein halbes Jahrhundert nach der Gründung der 'Econometric Society' eine solche Reife oder gar eine gewisse Vollendung erreicht, daß die Ökonometriker glauben, mit Stolz auf den Entwicklungsprozeß ihrer Wissenschaft zurückblicken zu können? Oder scheint es heute notwendig, mit Wehmut auf die hohe intellektuelle Leistung der ersten Ökonometriker sowie auf die schlichte Schönheit und den Realitätssinn der ersten ökonometrischen Beiträge hinzuweisen, weil die Ökonometrie inzwischen zu einer bloßen Technik verkommen ist, zu einer Technik, die zwar ständig neue Verästelungen und Verfeinerungen erfährt, die heute jedermann in unzähligen umfangreichen Programmpaketen zur Verfügung steht, die aber vor allem eingesetzt und weiterentwickelt wird, um Beifall in der akademischen Gesellschaft zu finden, aber nicht, um praktische Fragestellungen zu behandeln, weil man sich dann mit den vielen Widerhaken und Fußangeln der realen Welt abmühen müßte?"⁷

Die "Historisierung" der Ökonometrie wurde vor allem durch die vielbeachtete Arbeit von M. Morgan (1990) gefördert,⁸ der neben weiteren Analysen⁹ auch eine

⁴ So ist es mittlerweile keine Seltenheit mehr, daß Lehrbücher zu bestimmten Gebieten der Ökonometrie einen Umfang von 800 Seiten überschreiten, wie z. B. Davidson/MacKinnon (1993), Hamilton (1994) oder Hendry (1995). Die kaum zu bewältigende Fülle von behandelten Spezialproblemen findet eine Reihe spezieller Foren in Form von explizit "ökonometrischen" Zeitschriften, wie etwa (in Klammern jeweils das Gründungsjahr): "Econometrica" [1934], "Journal of Econometrics" [1973], "Econometric Reviews" [1982], "Econometric Theory" [1985], "Journal of Applied Econometrics" [1986]. Eine nützliche Übersicht über aktuelle Forschungsthemen bieten Oxley/George/Roberts (1995).

⁵ Tinbergen in Magnus/Morgan (1987, S. 136).

⁶ Vgl. dazu vor allem Kirchgässner (1983). Die Lehrbücher lassen diesen konzeptionellen Aspekt generell außer acht. Dies beklagen etwa Swamy/Conway/von zur Muehlen (1985, S. 4). Eine Ausnahme ist Hendry (1995).

⁷ Tewes (1991, S. 614) in einer Besprechung von Morgan (1990).

⁸ Vgl. etwa Tewes (1991), Heckman (1992) oder Neuberger (1995). Die drei Jahre zuvor erschienene Monographie von Epstein (1987) fand dagegen kaum Beachtung. Der Monographie Morgans waren zwei Studien vorausgegangen, Morgan (1987b) und Morgan (1988).

Reihe von Wiederabdrucken bedeutender Texte folgte. Neben der von Morgan zusammen mit Hendry herausgegebenen Sammlung¹⁰ ist hier vor allem die Reihe *The International Library of Critical Writings in Econometrics* zu erwähnen, in der bislang die für unseren Zusammenhang relevanten Bände *The History of Econometrics*, *Macroeconometric Modelling*, *The Methodology of Econometrics*, *Time Series* und *Bayesian Inference* erschienen sind.¹¹ Damit liegen nun eine Reihe von Vorarbeiten vor, auf die wir unsere Ausführungen aufbauen können, wenn auch aufgrund unseres Erkenntnisziels die Schwerpunkte zum Teil anders zu setzen sind.¹²

Wie in Teil A stellt sich auch hier zunächst die Frage des zeitlichen Beginns. Sicher hat es schon seit dem 17. Jahrhundert "ökonometrische" bzw. "empirische" oder "statistische" Untersuchungen im weitesten Sinne gegeben,¹³ doch wollen wir den Beginn mit den ersten vorrangig ökonomisch interessierten Untersuchungen über Konjunkturzyklen im 20. Jahrhundert mit den Arbeiten von Mitchell (1913), Moore (1914) und Persons (1919) ansetzen.¹⁴

In der Ökonometrie wurden methodologische Konzepte in noch geringerem Maße als in der theoretischen Statistik um ihrer selbst willen entwickelt. Insofern ist eine Einteilung in die Erkenntnisziele Deskription und Inferenz hier nicht mehr eindeutig möglich. Andererseits wurden statistische Untersuchungen in der Ökonomie erst durch die Arbeiten von drei Physikern, R. Frisch, T. C. Koopmans und J. Tinbergen, systematisch "stochastifiziert". Insofern erscheint es gerechtfertigt, die vorangehende Entwicklung als deskriptiv und explorativ zu bezeichnen.

⁹ Siehe etwa Morgan (1991), Morgan (1995b) und die in Nachfolge von Morgan (1990) entstandene Arbeit von Qin (1993), die auch zeitlich an Morgan anschließt. Die historische Entwicklung makroökonometrischer Modellbildung wird von Bodkin/Klein/Marvah (1991) behandelt.

¹⁰ Hendry/Morgan (1995a). Diese Sammlung enthält eine umfangreiche Einleitung, Hendry/Morgan (1995b), in der eine Reihe von Daten einer Reanalyse unterzogen werden.

¹¹ Darnell (1994a), Wallis (1994a), Poirier (1994a), Harvey (1994a), Polson/Tiao (1995a).

¹² Epstein (1987), dessen Arbeit sich zum Teil auf unveröffentlichtes Archivmaterial stützt, behandelt vorrangig die Entwicklung der "structural estimation" sowie kurz das Konzept der Exogenität und die von Sims propagierten vektorautoregressiven Modelle, geht aber nicht auf bayesianische Ansätze oder den Einfluß englischer Ökonometriker wie J. D. Sargan oder D. F. Hendry ein. Vgl. zu letzteren Gilbert (1989). Auch die in unserem Zusammenhang relevante inferenzstatistische Konzeption der Ökonometrie findet keine explizite Berücksichtigung, ebenso wenig wie der Bereich der Zeitreihenanalyse, obwohl Epstein betont, daß sie auf "our conception of econometrics" zunehmend Einfluß ausübe. Morgan (1990) untersucht verschiedene Aspekte: Die Entwicklung der statistischen Analyse von Konjunkturzyklen, die ökonometrische Entwicklung auf dem Gebiet der Nachfrageanalyse, Fehlermodelle und schließlich die auf T. Haavelmo zurückgehende 'probability revolution' der Ökonometrie.

¹³ Vgl. etwa Stigler (1954), Stone (1988) oder de Vauban (1994). Schumpeter (1965, S. 274ff) bezeichnete bereits die politischen Arithmetiker und Physiokraten als "Ökonometriker".

¹⁴ Epstein (1987, S. 12) setzt den Beginn der Ökonometrie mit Moore (1911) an: "When academic economists such as Marshall, Jevons and Menger became involved in policy debates they tended to concentrate on macroeconomic issues such as the currency system, where they could draw upon many years of actual experience to set up viable institutional frameworks. Effective empirical work in fact relied heavily on the skills of the historian and the jurist. Economic series were collected and plotted for trends but even at the turn of the century it was almost unknown to use such data to derive specific causal relationships."

Mit der Gründung der *Econometric Society* wurden diese deskriptiven und explorativen Ansätze programmatisch durch eine stärkere Verbindung von Wirtschaftstheorie, Mathematik und Statistik ersetzt. Hauptziel war nun, ökonomische Theorien zu testen und ökonomische Modelle soweit numerisch zu konkretisieren, daß sie für wirtschaftspolitische Eingriffe verwendet werden konnten. In methodologischer Hinsicht war dies mit verschiedenen Adaptionen der theoretischen (Inferenz-) Statistik verbunden, insbesondere seit Frischs Schüler T. Haavelmo die Neyman-Pearson-Theorie auch für die Ökonometrie als maßgebliche Sicht verkündete.

Das ursprüngliche Programm der Ökonometrie entpuppte sich spätestens seit dem Ende der siebziger Jahre als Illusion. Der bereits angesprochene Wandel von einer Theorienüberprüfung zur Datenanalyse ging zwar in methodischer Hinsicht mit einer breiten Auffächerung einher, dennoch haben sich seit Anfang der achtziger Jahre verschiedene Methodologien herausgebildet, die jeweils für sich beanspruchen, die aufgetretenen Probleme lösen zu können.

2. Deskription und Exploration

a. Allgemeines

Wenn man sich die zahlreichen Probleme vergegenwärtigt, die sich bei der Analyse ökonomischer im Vergleich zu naturwissenschaftlichen Daten *zusätzlich* stellen, so ist es zunächst erstaunlich, daß die dort entwickelten Verfahren hier überhaupt Anwendung fanden. Betrachtet man etwa den "Problemkatalog" O. Morgensterns, der allein für die Bestimmung so wichtiger Größen wie dem Volkseinkommen eine Fehlergröße von $\pm 15\%$ bis $\pm 20\%$ vermutete,¹⁵ so kommt der reinen Beschreibung sicher eine größere Bedeutung zu als in den Naturwissenschaften, zumal hier die Vorstellung kontrollierter und wiederholbarer Experimente kaum anwendbar ist: "Viele wirtschaftliche Beobachtungen sind einmalig und *nicht* wiederholbar. Gewöhnlich steht man geschichtlichen Vorgängen gegenüber."¹⁶

Unbestritten ist jedoch, daß zentrale Konzepte der Statistik übernommen wurden. Vor allem die Korrelation als deskriptives Maß für den Zusammenhang zwischen zwei oder mehreren Größen fand hier Anwendung, auch bei der in der Ökonomie besonders relevanten Analyse von Zeitreihen.¹⁷ Andererseits interessierten zunächst nicht nur die Zusammenhänge, sondern auch Verläufe "an sich". Die Analyse von Trends und Zyklen entwickelte sich zu einem vorrangigen Erkenntnisziel.

Drei verschiedene Ansätze stehen am Beginn der Entwicklung: erstens der stark auf historische Einmaligkeiten bezugnehmende Ansatz von W. C. Mitchell, zweitens der sehr viel stärker formale Methoden der Statistik adaptierende Ansatz von

¹⁵ Vgl. Morgenstern (1952, S. 12).

¹⁶ Ebda., S. 30. Hervorhebung im Original.

¹⁷ Auf die Entwicklung der Regressionsrechnung im Zusammenhang mit Nachfrageanalysen wollen wir hier nicht eingehen. Siehe dazu Christ (1985) und Morgan (1990, Kap. 5 und 6).

H. L. Moore und drittens (und am folgenreichsten) die Konstruktion der sogenannten Harvard-Kurven durch W. Persons.

In der Folgezeit wurden eine Reihe von Verfahren vorgeschlagen, um aus ökonomischen Reihen Trends und Zyklen zu isolieren, wobei immer wieder auf die Instabilität ökonomischer Verhältnisse hingewiesen wurde. Man versuchte, die von der Statistik bereitgestellten Verfahren so weiterzuentwickeln, daß diesem Aspekt Rechnung getragen werden konnte.

Fundamentale Bedeutung kam unmittelbar nach deren Erscheinen den in Abschnitt A 4 b besprochenen Arbeiten von Yule und Slutsky zu. Die von ihnen vorgestellten Mechanismen waren eine klare Alternative zu den traditionellen Komponentenmodellen. In den Komponentenmodellen war die Fehlergröße lediglich ein zusätzlicher, die anderen Komponenten überlagernder Störfaktor. Die von Yule und Slutsky präsentierten Modelle stellten dagegen die Störgröße in den Mittelpunkt. Sie war demnach die eigentliche Ursache für den Verlauf einer Reihe und auch für Erscheinungen wie Trends oder Zyklen. Hier wurde zwar ein Datengenerierungsmechanismus postuliert, es fanden aber keine Schätzungen von Parametern statt. Insofern handelt es sich auch bei diesen Modellen zunächst um Illustrationen bzw. Beschreibungen möglicher Mechanismen ohne inferenzstatistische Ambitionen. Nach der Einführung dieser Modelle trennte sich die weitere Entwicklung. Eine "traditionelle" Richtung lehnte die Implikationen dieser Modelle ab und vertrat weiterhin die traditionelle Komponentenvorstellung. Hauptvertreter dieser Richtung wurde W. C. Mitchell, der zunächst mit S. Kuznets, später mit A. F. Burns konsequent den von ihm bereits 1913 eingeschlagenen Weg weiterging. Er begründete 1946 damit eine "offizielle" Wirtschaftsstatistik der Messung von Konjunkturzyklen, die für Jahrzehnte die Arbeit des *National Bureau of Economic Research* bestimmen sollte, aber auch zu einer entschiedenen Ablehnung seitens der mittlerweile etablierten Ökonometrie führte, die ihm den Vorwurf des "Messens ohne Theorie" entgegenbrachte.

b. Trends, Zyklen, Korrelationen

Die Ausgangslage

In Abschnitt A 4 b hatten wir die Entwicklung der Zeitreihenanalyse und der Theorie stochastischer Prozesse von der statistischen Seite her angesprochen. Hier wollen wir uns nun der Entwicklung seitens der entstehenden quantitativ-empirischen Wirtschaftswissenschaft zuwenden.¹⁸

¹⁸ Eine Trennung beider Entwicklungen ist für die Zeit des beginnenden 20. Jahrhunderts sicher nicht exakt möglich. Dennoch erscheint uns die hier getroffene Einteilung legitim. Zum einen stehen bei den in diesem Kapitel zu besprechenden Arbeiten die inhaltlichen Probleme eher im Vordergrund als die methodischen, zum anderen sind die Verfasser hier Ökonomen und keine Statistiker oder Naturwissenschaftler. Ein weiteres Unterscheidungsmerkmal besteht darin, daß die nun folgenden Untersuchungen fast ausschließlich amerikanischer Herkunft sind, während die in Teil A beschriebene Entwicklung (ohne Ausnahme) kontinentalen Ursprungs war.

Am Beginn stand die empirische Analyse von Konjunkturzyklen. Zwar hatten schon C. Juglar und W. S. Jevons Zeitreihen hinsichtlich dieses Aspektes "statistisch" ausgewertet, doch erst mit den Arbeiten von drei "Pionieren", W. C. Mitchell, H. L. Moore und W. Persons, begannen systematische Untersuchungen, die die nachfolgende Entwicklung langfristig bestimmten. Alle drei gingen jedoch unterschiedliche Wege.

Mitchell publizierte 1913 ein über 600 Seiten umfassendes großformatiges Werk über *Business Cycles and their Causes*, in dem sich neben zahlreichen graphischen und tabellarischen Präsentationen auch der Versuch einer systematischen Auswertung des vorhandenen und zusammengestellten Materials für vier Länder und die Zeit von 1890 bis 1911 fand.¹⁹ Sein Ziel war die Überprüfung von Theorien, ausgehend von empirischem Material. In methodischer Hinsicht ging diese Arbeit kaum über Jevons und Juglar hinaus. Mitchell beschränkte sich auf graphische Vergleiche und die Berechnung von Mittelwerten für die beiden Jahrzehnte.²⁰ Seine Schlußfolgerungen waren überaus vorsichtig: Jeder Zyklus war seiner Ansicht nach hinsichtlich des Verlaufes einmalig und von der vorherigen Entwicklung sowie von externen Faktoren wie etwa dem Wetter oder politischen Ereignissen abhängig. Weiterhin vermutete er, daß die säkulare Änderung der Wirtschaftsorganisation einen Einfluß auf die einzelnen Zyklen ausübe und eine zusätzliche Variabilität in der wirtschaftlichen Entwicklung verursache.²¹

Eine systematische statistische Analyse, die nicht nur empirische Regelmäßigkeiten feststellen, sondern auch Kausalmodelle aufstellen wollte, wurde 1914 von Moore vorgestellt.²² Moore gelangte zu der Überzeugung, daß der wichtigste, anhaltende Grund für ökonomische Zyklen in den zyklischen Schwankungen der Ernteerträge zu finden sei, diese wiederum durch die zyklischen Witterungsbedingungen, den "rhythmically changing amount of rainfall" (S. 149) verursacht werde. Hierbei handelte es sich keineswegs um eine neue Entdeckung; neu war dagegen die von ihm angewandte statistische Vorgehensweise.²³ Er forderte von einer sinnvollen statistischen Methode, daß sie die Daten so weit wie möglich repräsentieren, die Beschreibung einer kausalen Beziehung ermöglichen und die wesentli-

¹⁹ Wesley Clair Mitchell [1874-1948] war von 1901 bis 1902 an der Universität von Chicago tätig, 1903 bis 1913 an der *University of California*, von 1913 bis 1919 und 1922 bis 1944 an der *Columbia University professor of economics*, von 1919 bis 1931 Direktor der *New School of Social Research* und einer der maßgeblichen Mitbegründer des *National Bureau of Economic Research* (1920).

²⁰ Vgl. Mitchell (1913, passim).

²¹ Vgl. auch Morgan (1990, S. 46f).

²² Moore (1914). Für Vorläufer vgl. Yule (1915, S. 303f) oder Nerlove/Grether/Carvalho (1979, S. 4f). Moores Arbeiten werden von Morgan (1990, S. 27) als erste "ökonometrische" Beiträge bezeichnet. Henry Ludwell Moore [1869-1958] promovierte 1896 an der *Johns Hopkins University* mit einer Arbeit über die Thünensche Theorie natürlicher Löhne, war von 1897 bis 1902 *Professor of Political Economy* am *Smith College*, Northampton, bis zu seiner vorzeitigen Emeritierung 1929 an der *Columbia University*. Von 1909 bis 1913 besuchte er Kurse in mathematischer Statistik und Korrelationstheorie bei Karl Pearson.

²³ Vgl. Moore (1914, S. 137f). Nach Yule (1915, S. 303) war die Arbeit bis dato "[...] from the statistical side, by far the most detailed discussion of the subject that has yet appeared." Siehe auch Qin (1993, S. 42), Morgan (1990, S. 26ff).

chen Charakteristika vorhandener Periodizitäten herausstellen sollte. Diese Forderungen sah er in der bislang noch nicht im ökonomischen Zusammenhang verwendeten Periodogramm- verbunden mit einer herkömmlichen Korrelationsanalyse eingelöst. Moore untersuchte zunächst die zyklischen Charakteristika des Wetters (genauer des Regenfalls als bestimmendem Faktor) anhand einer Periodogrammanalyse von Daten für das Tal von Ohio für die Jahre von 1839 bis 1910; dabei fand er periodische Schwankungen von 33 und 8 Jahren. Diese Ergebnisse korrespondierten in etwa mit dem von ihm ebenfalls untersuchten Regenfall in Illinois von 1870 bis 1910. Eine anschließend durchgeführte Korrelationsanalyse ergab für die Frühlings- und Sommermonate starke Zusammenhänge zwischen diesen Daten auf der einen und den Erträgen landwirtschaftlicher Produkte (Mais, Kartoffeln etc.) auf der anderen Seite. Anschließend untersuchte er die Korrelation zwischen den Erträgen und Preisen dieser Produkte, wobei er die jährlichen prozentualen Veränderungen (Wachstumsraten) der Durchschnittspreise und der gesamten US-Produktion zugrundelegte. Eine Regressionsrechnung der Produktions- und Preisänderungen brachte dabei für alle untersuchten Produkte hohe Korrelationen. Ähnliche Ergebnisse lieferten Berechnungen mit den Ertragsgrößen anstelle der Produktion sowie eine Untersuchung der Beziehung zwischen industrieller Produktion und Preisen. Für die Untersuchung der Zyklen wurden dagegen keine Wachstumsraten, sondern Abweichungen von linearen Trends zugrundegelegt. In methodischer Hinsicht ging diese Arbeit weit über die Untersuchung Mitchells hinaus. Es ist daher nicht erstaunlich, daß sie auch seitens der Statistik Beachtung fand und von G. U. Yule ausführlich besprochen wurde. Yule kritisierte zu Recht, daß Moore die Periodogrammanalyse lediglich auf die Wetter-, nicht aber auf die Ertrags-, Produktions- und Preisdaten angewandt hatte.²⁴ Außerdem sei die Eliminierung des Trends durch eine *lineare* Funktion nicht angemessen:

“What right have we, in the first place, to assume that over so long an interval as forty years the secular trend is linear and that the deviation may be regarded as a periodic movement? [...] Would it not be better to assume, as would be more natural, that the secular trend may be more closely represented by a curve? If this were done it might be necessary to confine the study of the periodic movements to those of short period (eight years, or anywhere between seven and eleven years, to take the figures of various writers); but after all, when most people write of the ‘trade cycle’ they are not thinking of any such period as thirty-three years, *which does not strike the eye on the graph*, but of those short period waves which form so conspicuous a feature of the curves of foreign trade, the marriage rate, prices, unemployment and so forth.”²⁵

Trotz dieser Kritik sah Yule in der Arbeit einen wichtigen Beitrag. Wie wir in Abschnitt A 4 c gesehen hatten, hat er selbst diese Problematik später systematisch aufgegriffen und damit die angewandte statistische Zeitreihenanalyse begründet.²⁶

²⁴ Moore (1914, S. 44) begründete dies mit dem hohen Rechenaufwand.

²⁵ Yule (1915, S. 305). Hervorhebung von uns.

²⁶ Moore hat 1923 eine erweiterte Fassung seines Ansatzes vorgelegt, in der er Yules Kritik aufnahm und auch die ökonomischen Daten in die Periodogrammanalyse miteinbezog. Darin postulierte er nun die Bahn der Venus, die sich alle 8 Jahre zwischen Sonne und Erden stellt, als letzte Ursache für zyklische Witterungsänderungen gleicher Dauer.

Den dritten Ausgangspunkt bildeten die Studien von W. Persons. Bereits 1910 ging W. Persons auf die Arbeiten von Hooker und Yule ein,²⁷ zeigte weitere Anwendungsmöglichkeiten und schlug zusätzlich vor, die langfristige Tendenz ("long-time tendency") durch eine Funktion auszuschalten, wenn a priori Gründe vorhanden seien, die die Existenz einer solchen Funktion in den zugrundeliegenden Daten nahelegten. So zeige etwa die Wachstumsfunktion $y = bc^x$ eine sehr angemessene Anpassung an die Entwicklung der amerikanischen Bevölkerung zwischen 1850 und 1900.²⁸

Später sah er diese Frage pragmatischer.²⁹ Persons hatte vom *Harvard Committee for Economic Research* die Aufgabe erhalten, die verbreiteten Ad-hoc-Verfahren zur Vorhersage zukünftiger geschäftlicher Entwicklungen zu systematisieren und methodisch zu verbessern. Zu diesem Zweck ging er davon aus, daß ökonomische Reihen sich aus Komponenten zusammensetzten, die unterschiedlichen Wirkungskräften ausgesetzt waren: säkulare, saisonale, zyklische sowie irreguläre Kräfte.³⁰ Um zyklische Bewegungen in ihrer Bedeutung würdigen zu können, müsse zunächst der Trend ausgeschaltet werden. Ein "Ausprobieren" anhand verschiedener Reihen ergab, daß die Anpassung von Trendfunktionen zufriedenstellendere Resultate lieferte als die verbreitete Methode der gleitenden Mittelwerte. Problematisch war jedoch die geeignete Wahl der zugrundeliegenden Periode: Hier könne die Mathematik ebenso wenig Hilfestellung leisten wie bei der genauen Wahl der Trendfunktion. Die Untersuchung der Zyklen³¹ gestalte sich als besonders schwierig, da diese nicht von der irregulären Komponente zu trennen seien.³² Insofern erweise sich eine Glättung anhand eines gleitenden Mittelwertes als praktikabel. Um Zyklen unterschiedlicher Reihen miteinander zu vergleichen, solle man diese Komponente "normalisieren", d. h. die prozentualen Abweichungen vom Trend betrachten. In einem zweiten Schritt konstruierte Persons einen Index, der – mit den beschriebenen Methoden ausgewertet – möglichst gute Prognosen erlauben sollte.³³ Hier ging er auf ungewöhnliche Weise vor:³⁴ Er suchte aus 50 Datenreihen 20 heraus, die ähnliche Perioden, aber unterschiedliche Phasen (Hoch- und Tiefpunkte) aufwiesen. Um jede mit jeder hinsichtlich ähnlicher Verlaufsmuster, die möglicherweise zeitversetzt in Erscheinung traten, zu vergleichen, mußte er 190 Vergleiche durchführen. Für diese Vergleiche wählte er drei "Beobachter", denen jeweils zwei Reihen mit unterschiedlichen Verschiebungen auf eine Leinwand projiziert wurden. Diese Beobachter mußten dann notieren, ob sich ein hoher,

²⁷ Persons (1910). Zu den Arbeiten von Yule und Hooker siehe oben, Abschnitt A 4 b. Warren M. Persons [1878-1937] studierte an der Universität von Wisconsin, ging 1919 an die *Harvard University*, verließ diese jedoch bereits 1928 und arbeitete seither als Wirtschaftsberater.

²⁸ Siehe die Abbildung in Persons (1910, S. 317).

²⁹ Vgl. zum folgenden auch Qin (1993, S. 42), Morgan (1990, S. 56-70).

³⁰ Persons (1919a).

³¹ Auf die Behandlung der saisonalen Komponente, die sich vergleichsweise unproblematisch gestaltete, wollen wir hier nicht näher eingehen.

³² Persons (1919a, S. 33).

³³ Vgl. Persons (1919b).

³⁴ Vgl. Morgan (1990, S. 61).

mittlerer oder niedriger Zusammenhang zeigte und – im Falle eines Zusammenhanges – um wie viele Monate verschoben dieser auftrat. Für die vielversprechendsten Ergebnisse führte er Korrelationsrechnungen für verschiedene “lags” durch. Aus den bei diesen Berechnungen am meisten übereinstimmenden Reihen wurden dann schließlich drei Indizes konstruiert: die sogenannte Harvard A-B-C Kurven, die zukünftig für Prognosen verwendet werden konnten.

Persons’ Methoden waren offensichtlich durch eine Reihe von heuristischen Hilfskonstruktionen gekennzeichnet, dennoch bestimmten sie für lange Zeit das Bild der statistischen Ökonomie. Seine Harvard-Kurven bildeten die Grundlage der meisten Konjunkturprognosen.³⁵

In der Folgezeit wurden die von Mitchell, Moore und Persons angesprochenen Themen weiter ausgebaut. Dabei spielte vorrangig die univariate Analyse der Reihen eine Rolle, insbesondere eine Aufteilung in die Komponenten Trend und Zyklus sowie eine irreguläre Komponente, denen jeweils unabhängige verursachende Faktoren unterstellt wurden.³⁶ Diese Vorstellung wurde durch die Arbeiten von Yule und Slutsky 1927 in Frage gestellt. Die Zufallsgrößen waren demnach nicht mehr das lästige oder uninteressante fünfte Rad am Wagen, sondern der Motor der Entwicklung. Nicht alle wollten (oder konnten) dieser folgenreichen Interpretation folgen; zu diesem Zeitpunkt trennte sich daher die weitere Entwicklung. Zum einen wurde der “traditionelle” Komponentenansatz weiter verfolgt, vor allem durch J. A. Schumpeters *Business Cycles* sowie A. Burns’ und W. C. Mitchells *Measuring Business Cycles*, zum anderen wurde eine “ökonometrische” Richtung geboren, die die statistische Analyse ökonomischer Daten rigoros in einen stochastischen – und vor allem multivariaten – Rahmen setzte und der hier geschilderten Entwicklung ein theorieloses Messen und damit unwissenschaftliches Vorgehen vorwarf.³⁷

Insgesamt überwogen in dieser Phase univariate Analysestrategien, doch blieb natürlich die Frage nach Zusammenhängen zwischen Reihen nach wie vor relevant. Auch in diesem Kontext zeigte sich ein Problem, das schon bei der Untersuchung von Trends und Zyklen auftauchte: die Frage nach der zeitlichen Konstanz der strukturellen Beziehungen.

Wir wollen im folgenden als erste Orientierung einen Überblick über diese zeitreihenanalytischen Diskussionen geben, die vor allem in den Zeitschriften *Journal of the American Statistical Association* und *Review of Economic Statistics* thematisiert wurden.³⁸

³⁵ Nach Nerlove/Grether/Carvalho (1979, S. 12) ist das steigende Interesse an der Zeitreihenanalyse in den zwanziger Jahren vor allem den Arbeiten Persons’ zu verdanken. Sie betonen jedoch, daß zu diesem Zeitpunkt die Auffassung von Zeitreihen als Kompositum verschiedener Komponenten sich bereits allgemein durchgesetzt hatte: “[...] the basic ideas about the meaning of the components were fixed at a relatively early date. Later work refined and systematized these notions, but much of the conceptual content of this literature predates the modern period.” Sie betonen auch, daß die spätere Entwicklung sich stärker auf die Methoden selbst, als auf eine Begründung für ihre Anwendung konzentriert habe (S. 10).

³⁶ Bei unterjährigen Reihen wurde zusätzlich eine saisonale Komponente berücksichtigt.

³⁷ Siehe dazu den nächsten Abschnitt. “Ökonometrische” Traditionen beginnen etwa Anfang der dreißiger Jahre, während wir hier die Entwicklung bis 1946 verfolgen.

Der Trend

Ein wichtiges (das erste) Problem, das sich in bezug auf methodologische Probleme im Zusammenhang mit der Untersuchung ökonomischer Entwicklungen stellte, war verständlicherweise die (statistische) Behandlung des Trends. Die im Mittelpunkt des Interesses stehenden ökonomischen Verläufe der unmittelbar zurückliegenden Vergangenheit bis hin zur Gegenwart waren in der Regel durch stark ausgeprägte Trendentwicklungen gekennzeichnet. Zwar richtete sich das Interesse vorrangig auf die Analyse einer konjunkturellen, möglicherweise beeinflussbaren "Komponente", die als eine sich in der Regel unabhängig vom Trendverlauf konstituierende Entwicklung gesehen wurde, doch finden sich auch Untersuchungen, die den Trendverlauf selbst einer genaueren Analyse zugänglich machen wollten. Hierzu gab es unterschiedliche Ansätze. Eine weitverbreitete Ansicht war, daß sich die meisten ökonomischen Entwicklungen säkular linear vollziehen, wie etwa Crum 1922 feststellte: "It is customary in the analysis of ordered statistical series of the historical economic sort to eliminate the growth factor, *assumed constant*, by fitting a straight line to the original data."³⁹ Dieser Trend werde üblicherweise durch die Kleinstquadratmethode angepaßt, die die Summe der quadrierten *Abweichungen* minimiere. Crum war nun der Ansicht, daß nicht die quadrierten Abweichungen, sondern die Summe der quadrierten *prozentualen* Abweichungen minimiert werden sollte. Diese Methode sei darüber hinaus besonders geeignet, die zyklische Komponente einer Reihe angemessen zu isolieren.⁴⁰ Crums Ansatz stieß unmittelbar darauf auf Widerspruch. Noch im selben Jahrgang findet sich eine Notiz von Holbrook Working. Working wies darauf hin, daß die von Crum vorgeschlagene Methode zunächst einmal einen wesentlich größeren Rechenaufwand bedeute als die übliche Minimierung der Summe der gewöhnlichen Abweichungen und ein derartiger größerer Aufwand nur durch einen erheb-

³⁸ Dies ist insofern gerechtfertigt, weil das *Journal of the American Statistical Association* das zentrale Organ war, in dem *statistische* Methodenfragen diskutiert wurden. Die Zeitschrift *Review of Economic Statistics* wurde 1919 vom *Harvard Committee for Economic Research* ins Leben gerufen und setzte sich das Ziel, eine bessere Grundlage ökonomischer Daten zu schaffen sowie "modern methods of statistical analysis which have hitherto been utilized more extensively in other sciences than in economics" in die Ökonomie einzuführen. Zitiert nach Morgan (1990, S. 58) aus dem Vorwort der ersten Ausgabe. Auf zwei für dieses Thema bedeutsame Entwicklungen der zwanziger Jahre können wir vorerst nicht eingehen: die deutsche empirische Konjunkturforschung (A. Spiethoff, E. Altschul, E. Wagemann) und die kontinentale Schule der mathematischen Statistik, insbesondere O. Anderson. Gerade letzterer hat nicht nur Beiträge über die Korrelation von Zeitreihen sowie zur *variate-difference*-Methode geliefert, die auch in der anglo-amerikanischen Forschung beachtet wurden, sondern diese auch mit einem dezidierten stochastischen Konzept verbunden. Bislang liegt keine systematische Untersuchung insbesondere hinsichtlich der Beziehung zwischen der empirischen Konjunkturforschung und der kontinentalen Schule, die sich z. B. in der Person Oskar Andersons manifestiert, vor. Dies muß einer späteren Analyse vorbehalten bleiben. Morgan (1990) beschränkt sich in ihrer Darstellung auf die zyklische Dimension und wählt zudem eine Systematisierung, die wir nicht für sinnvoll halten. Für eine Reihe einzelner Aspekte können wir jedoch auf ihre Arbeit zurückgreifen.

³⁹ Crum (1922, S. 210). Hervorhebungen von uns.

⁴⁰ Vgl. ebda., S. 215. Crum präsentierte diesen Ansatz 1921 auf einer Tagung der *American Mathematical Society*.

lichen Zuwachs an Erkenntnisgewinn oder durch bedeutende Unterschiede in den errechneten Ergebnissen zu rechtfertigen sei. Working wendete die von Crum vorgeschlagene Methode auf die Entwicklung der jährlichen Produktion verschiedener Getreidesorten von 1879 bis 1913 an und gelangte dabei zu dem Schluß, daß die Unterschiede der mittels beider Methoden erzielten Ergebnisse so gering ausfallen, daß der erhöhte Rechenaufwand für die Probleme, die sich in der Praxis stellen, nicht gerechtfertigt sei: "Professor Crum's criticism of the usual method is theoretically valid, but of little practical significance."⁴¹

Eine Alternative zur Berechnung eines linearen Trends schlug zwei Jahre später W. I. King vor. King betonte, daß die mathematische Analyse weder Faktenwissen noch logische Überlegungen ersetzen dürfe. So sei es durchaus möglich, daß ein Trend durch ein ungewöhnliches Ereignis einen Bruch aufweise, wie dies z. B. bei der Preisentwicklung von 1900 bis 1924 ersichtlich sei: Vor 1914 sei der Trend wesentlich durch die sinkenden Produktionskosten für Gold bestimmt gewesen, nach dieser Zeit jedoch durch die Papiergeldinflation und die Geschwindigkeit, mit der die Regierung Papiergeld produzieren konnte, beeinflußt worden.⁴² In solchen Fällen sah es King als sinnvoll und notwendig an, separate Trends anzupassen.

Mit welcher Methode sollte dies geschehen? Hier könne man per "judgement" oder streng mathematisch vorgehen. Die Wahl einer mathematischen Funktion sei jedoch zu begründen: "The fact should always be kept in mind that the use of a mathematical curve implies a belief that the force or forces dominating the variable act according to a mathematical law."⁴³ Nur in diesem Fall sei eine Anpassung einer mathematischen Funktion auch angemessen.⁴⁴ Falls man sich für eine mathematische Funktion entschlöße: Welche Form sollte diese haben? Dies müsse anhand einer visuellen Inspektion der Angemessenheit der angepaßten Funktion geklärt werden. Wenn man manuell einen Trend einzeichnen könne, der den Verlauf besser wiedergäbe als die Funktion, so sei nicht davon auszugehen, daß die Funktion den wahren Trend widerspiegele. Ein solcher Vorschlag sei sicher provokativ, aber er berücksichtige eher als andere die Realität:

"The fact remains, however, that even the most skilled mathematicians usually decide by observation whether or not a mathematical formula is giving satisfactory results. In practice, they, as well as the rest of us, are inclined to believe that judgement is after all superior to any rule of thumb procedure."⁴⁵

⁴¹ Working (1922, S. 502).

⁴² Vgl. King (1924, S. 470).

⁴³ Ebda.

⁴⁴ Eine Alternative sei die Bildung des Anteils an einer anderen Variable, die einen ähnlichen Verlauf aufweise und mit der untersuchten in Verbindung stehe, wie etwa die Weizenproduktion und die Bevölkerungszahl oder die Anzahl der Bankrotte und die Gesamtzahl der Unternehmen. Wenn eine derart konstruierte Reihe noch einen Trend aufweise, seien die obigen Überlegungen anzuwenden.

⁴⁵ King (1924, S. 471).

Die Anpassung mittels der Methode der kleinsten Quadrate führe nicht zwingend zu optimalen Resultaten. King führte ein anschauliches Beispiel an. Eine Trendberechnung nach der Methode der kleinsten Quadrate für fiktive Daten ergäbe die folgende Regressionslinie

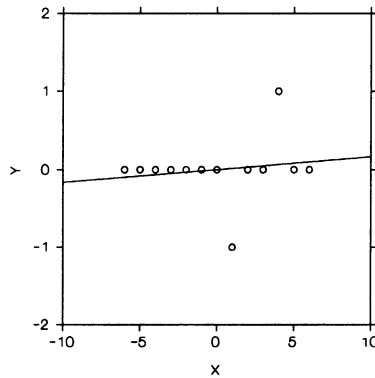


Abb. 5: "Ausreißer"-Illustration von King (1924, S. 472). Graphisch umgesetzt.

Ein idealer Trend sei derjenige, der die "dominant force" wiedergäbe. Offensichtlich sei hier eine waagerechte Linie angemessener, der Trend weise jedoch aufgrund zweier abweichender Werte (in moderner Notation: "Ausreißer") eine Steigung auf. Als Alternative schlug King einen gleitenden Median vor, der von stark abweichenden Werten nicht beeinflusst werde.⁴⁶ Anhand eines solchen gleitenden Medians könne dann ein Trend angepaßt werden.

Der von King vorgeschlagene erste Schritt, den Trend anhand einer korrespondierenden Reihe zu eliminieren, wurde auch von Hall (1925) befürwortet, falls zwei Reihen ähnlich und homogen seien. In solch einem Fall sei dieses Vorgehen angemessen, denn: "Since the two series are similar and homogeneous, they are assumed to have approximately the same secular trends and therefore their ratios will contain little or no element of secular trend but only indications of cyclical fluctuations [...]."⁴⁷ Diese Methode könne jedoch nur angewendet werden, wenn diese "Ähnlichkeit" des Trendverlaufes Bestand habe. Da aber andererseits ein Verfahren möglichst allgemein sein und nicht von anderen Reihen abhängen sollte, sei nach einer weiteren Methode Ausschau zu halten.⁴⁸ Hall schlug daher einen "moving secular trend" vor. Die üblicherweise vorgenommene Berechnung eines linearen Trends werde in der Regel aus rechentechnischen, nicht aus inhaltlichen

⁴⁶ Vgl. ebda., S. 474.

⁴⁷ Hall (1925, S. 13). In der modernen Konzeption der "Kointegration" wäre dies ein Spezialfall für $y_t = \theta x_t + \varepsilon_t$ mit $\theta_t = 1$.

⁴⁸ Hier stand eindeutig die univariate Analyse im Mittelpunkt. Die Elimination des Trends anhand einer zweiten Reihe sollte nicht einer Untersuchung der Beziehung beider Größen dienen, sondern lediglich eine "Bereinigung" bewirken, um dann die untersuchte Reihe weiter (univariat) auszuwerten.

Überlegungen durchgeführt, denn: "it is improbable that the secular trend is strictly a straight line because it is unlikely that any economic phenomena will be encountered which is growing or increasing at a constant rate."⁴⁹ Plausibler sei, daß der Trend eine Funktion höherer Ordnung darstelle, jedoch nicht für die gesamte betrachtete Periode, sondern "represented only by a group of functions placed end to end and which have no necessary relation to each other, and which are the result of the course of economic events producing the series under observation."⁵⁰

Halls Vorschlag lautete: Anstelle einer einzigen, den gesamten betrachteten Zeitraum umfassenden Funktion solle man für jeweils eine geringe Anzahl von Werten (z. B. 4) einen linearen Trend berechnen, den Funktionswert des letzten Zeitpunktes als Wert für den "moving trend" annehmen und dann die gleiche Prozedur, jeweils um einen Zeitpunkt verschoben, bis zum Ende der Reihe wiederholen.⁵¹

Überblickt man die Arbeiten dieser Periode hinsichtlich der Bestimmung des Trends, so läßt sich konstatieren, daß man das Problem der Trendbestimmung klar erkannte, aber keinen eindeutigen Lösungsvorschlag liefern konnte. Als E. Frickey 1934 das Problem noch einmal aufgriff, verglich er beispielhaft anhand der Eisenproduktion der Vereinigten Staaten verschiedene Trendanpassungen. Sein Überblick brachte nicht weniger als 23 verschiedene, in der Literatur vorgeschlagene Trendanpassungen für diese Kurve zutage, wobei die jeweilige Wahl auch erheblichen Einfluß auf die resultierende Zyklenlänge der "Rest"komponente hatte: diese variierte von drei bis über fünfzig Jahre.⁵²

Die zyklische Komponente

In Abschnitt A 4 b hatten wir bereits die statistische Entwicklung der Analyse zyklischer Erscheinungen anhand des Periodogramms erwähnt. Starkes Aufsehen im Bereich der Konjunkturzyklenanalyse erregte nach den Untersuchungen Moores 1922 die Arbeit von W. H. Beveridge: eine Periodogrammanalyse eines anhand von Notierungen aus 40 Orten konstruierten europäischen Weizenpreisindex von 1500 bis 1869 sowie von Niederschlagsdaten der Jahre 1850 bis 1921.⁵³ Beveridge

⁴⁹ Hall (1925, S. 14. Hervorhebung von uns.

⁵⁰ Ebda.

⁵¹ Diese Technik entspricht der Berechnung eines gleitenden Mittelwertes, wobei hier nicht der mittlere, sondern der letzte berechnete Wert als "Trend" wert verwendet wird. Der Grund hierfür ist der Zweck der Berechnungen: sie sollen vorrangig der Prognose dienen. Siehe zum Zusammenhang dieser Verfahren Thome (1992a).

⁵² Vgl. Frickey (1934).

⁵³ Diese Arbeit hätte unter Umständen auch bereits in Teil A erwähnt werden können. Da sie jedoch 1. keine eigenen Methoden entwickelte, 2. das Anwendungsgebiet ökonomische Reihen sind und 3. die Arbeit vorrangig von der Ökonomie und nicht mehr von der theoretischen Statistik rezipiert wurde, ist ihre Erwähnung an dieser Stelle angemessener. Lord [seit 1946] William Henry Beveridge [1879-1963] arbeitete nach dem Studium in Oxford zunächst von 1902 bis 1909 als *Stowell Civil Law Fellow* am *University College*, Oxford und wurde anschließend Leiter der Arbeitsabteilung des *Board of Trade*, wo er ein Arbeitsvermittlungssystem und eine Arbeitslosenversicherung begründete, anschließend war er für das Verteidigungs- und Ernährungsministerium tätig. 1919 wurde er Direktor der *London School of Economics and Political Science*, das er zu einer weltweit anerkannten Institution ausbaute. Hier wurde er auch Leiter des 1929/30 von der Laura-Spelmann-Rok-

fand in der Weizenpreisreihe (nach immensem Rechenaufwand) nicht weniger als 19 bedeutsame Zyklen, die er auch verschiedenen Tests unterzog. Sein Ziel war vorrangig eine Untersuchung meteorologischer Zyklen. Da hierfür aber nicht ausreichend lang zurückreichende Reihen vorhanden waren, mußte statt dessen eine Reihe herangezogen werden, die mit jener in unmittelbarem Zusammenhang stand, aber weiter zurückverfolgt werden konnte: dies waren historische Aufzeichnungen von Getreidepreisen. Seine Ergebnisse waren seiner Ansicht nach eindeutig:

“The agreement between the two curves is so striking that it could not possibly be due to chance. The test shows the general correctness of the hypothesis [...], namely that some or all of the periods shown by analysis of wheat prices from 1545 to about 1850 are real, and have been operative as rainfall cycles from 1850 to 1921.”⁵⁴

Beveridges Analyse sowie die in Abschnitt A 4 b genannten Periodogramm-Arbeiten, die zum Teil auch ökonomisches Material verwendeten, wurden von der statistischen Ökonomie zum Teil weitergeführt, doch sah man schon früh die Grenzen dieses Ansatzes. Bereits mit einer großen Portion Skepsis war die Untersuchung von Crum (1923) ausgestattet. Crum untersuchte monatliche New Yorker Zinsdaten von 1874 bis 1913, für die er, unter anderem durch die Verwendung der Periodogrammanalyse, einen Zyklus mit einer durchschnittlichen Länge von 40 Monaten feststellte. Die Anwendung dieses Verfahrens, die er für naturwissenschaftliche Phänomene durchaus als angemessen ansah, hielt er in bezug auf ökonomische Daten für wenig erhellend. Diese wiesen Zyklen ungleicher Länge und Amplitude auf, weiterhin seien ökonomische Zyklen in der Regel asymmetrisch⁵⁵ und variierten ihre Form. Schließlich sei zum einen von einem Strukturwandel ökonomischer Phänomene, die langfristig betrachtet werden, auszugehen, zum anderen würde die Struktur auch immer wieder durch außergewöhnliche Ereignisse beeinflusst, die unter Umständen auch nachhaltige Strukturveränderungen bewirken könnten.⁵⁶ Crums Fazit lautete:

“[...] it is very doubtful whether the attempt to analyze the form of the cycle by the Fourier method will assist in describing or understanding the cycle. The average form of the cycle is at best not very precisely determined, and its breaking up into a group of harmonic components of differing periods and amplitudes is certainly bewildering and probably furnishes no true insight into the nature of economic fluctuation.”⁵⁷

kefeller-Stiftung finanzierten Internationalen Wissenschaftlichen Komitees für die Erforschung der Preise. Nach kurzer Rückkehr in Oxford seit 1937 verfaßte er 1942 während diverser Tätigkeiten für die Regierung den weltberühmten “Beveridge-Report” und gilt als Begründer des modernen britischen Wohlfahrtsstaates. Von 1941 bis 1943 war er Präsident der *Royal Statistical Society*.

⁵⁴ Beveridge (1922, S. 449).

⁵⁵ Das war bereits Beveridge (1922) aufgefallen.

⁵⁶ Vgl. Crum (1923, S. 18, 24, 25).

⁵⁷ Ebda., S. 17.

An dessen Stelle sollten daher andere, eher informelle Verfahren treten, wobei er die Bedeutung einer graphischen Darstellung der Originalreihe betonte.⁵⁸ Der von Crum trotz aller methodischen Vorbehalte dennoch postulierte Zyklus von etwa 40 Monaten wurde im selben Jahr von Kitchin bestätigt. Kitchin untersuchte für dieselbe Epoche weitaus umfangreicheres Material: vor allem clearings, Preise und Zinsen für die USA und Großbritannien.

Kitchin verwendete eher heuristische Verfahren, wie etwa die Berechnung der durchschnittlichen Abstände von Maxima und Minima in den einzelnen Reihen sowie Verbindungslinien der Maxima und Minima in graphischen Darstellungen der verschiedenen Zeitreihen, die jeweils etwa die 40-Monats-Perioden trennen (Abb. 6).

Wenn auch diese graphische "Evidenz" nicht unbedingt jeden von der Existenz dieses Zyklus überzeugen dürfte, so haben sie ihrem "Entdecker" dennoch zu unverhofftem Ruhm verholfen: Seit Schumpeter (1939) tragen Zyklen dieser Länge den Namen "Kitchin-Zyklen".

Die Periodogrammanalyse genoß bereits zu diesem Zeitpunkt kein hohes Ansehen mehr in statistischen Untersuchungen ökonomischer Daten und dies nicht, weil ihre Anwendung zu rechenaufwendig war oder hohe mathematische Kenntnisse voraussetzte, sondern weil man bereits klar ihre beschränkte Anwendungsmöglichkeit auf ökonomische Phänomene erkannt hatte.⁵⁹ Dies zeigte sich auch in der 1927 erschienenen zweiten großen Untersuchung von Mitchell, die sich als Fortsetzung und Erweiterung seines ersten Werkes verstand.⁶⁰ Er verteidigte darin seinen Ansatz von 1913 und baute ihn systematisch aus.⁶¹

Nach einer vergleichenden Übersicht über verschiedene Konjunkturtheorien stellte Mitchell fest, daß die Vielzahl der existierenden Theorien eine quantitative Analyse erfordere.⁶² Andererseits sprach er sich auch für eine historische Perspektive⁶³ sowie für eine letztendlich pragmatische Vorgehensweise aus: "Wir wollen Einsicht in die wirklichen Verhältnisse, die Theorien interessieren uns nur, soweit sie uns dazu verhelfen" (S. 57). In methodisch-statistischer Hinsicht beklagte er dagegen eine bislang unzureichende Entwicklung.⁶⁴ Insbesondere der Trend sei in der Regel lediglich "ausgeschaltet" worden, um Konjunkturzyklen zu isolieren, wobei dann angenommen werde, daß diese sich unabhängig vom Trendverlauf entwickeln – eine seiner Ansicht nach unrealistische Vorstellung. Unabhängig von der konkreten Berechnungsart des Trends (nach der Methode der kleinsten Quadrate, als gleiten-

⁵⁸ "It is only from a chart of the cycles that one can tell with assurance whether the irregular disturbances are so important as to be likely to invalidate the results of the periodogram analysis." Ebda., S. 26.

⁵⁹ Insofern ist die in der Forschung immer wieder behauptete "euphorische" Bezugnahme auf das Periodogramm in dieser Phase zu relativieren.

⁶⁰ Mitchell (1927 [1931]). Das Buch erschien 1931 in einer von E. Altschul herausgegebenen, vom Verfasser durchgesehenen und ergänzten deutschen Übersetzung.

⁶¹ Zu dieser Zeit galt W. Persons bereits als Autorität. Er ist in Mitchell (1927) der meistzitierte Autor.

⁶² Vgl. Mitchell (1927 [1931, S. 52f]).

⁶³ Vgl. ebda., Kap. I, V 5: "Theorie und Geschichte".

⁶⁴ Vgl. ebda., S. 180.

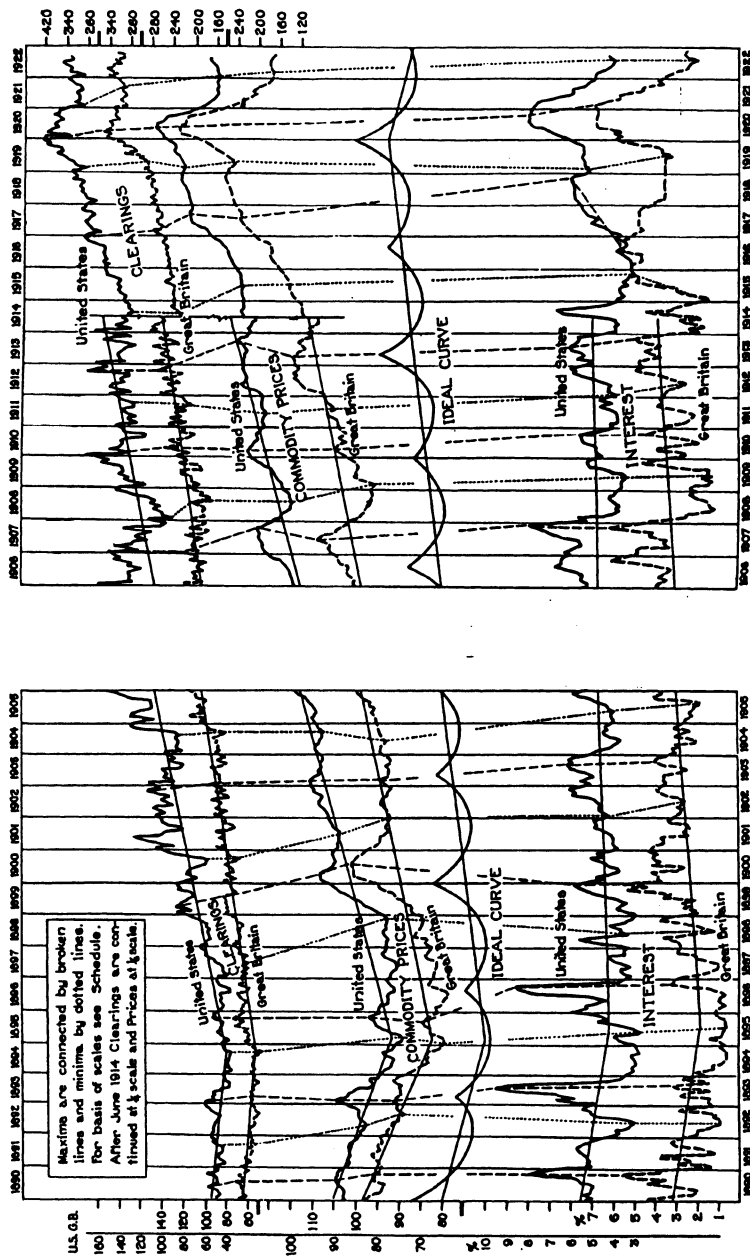


Abb. 6: Graphische Darstellung der "Kitchin"-Zyklen. Aus: Kitchin (1923, S. 12f).

der arithmetischer Mittelwert oder Median, durch Verhältnisbildung) bleibe die Entscheidung über die Angemessenheit der Anpassung letztendlich dem Forscher überlassen: "[...] die Literatur beweist, daß der eine einen Trend gut findet, der einem anderen nicht genügt. Das persönliche Urteil fällt dabei entscheidend in die Waagschale."⁶⁵ Hinsichtlich der Analyse von Konjunkturschwankungen mittels der Periodogrammanalyse äußerte Mitchell ähnliche Bedenken wie bereits Crum (1922).⁶⁶ Die letzte Komponente schließlich, die Zufallsschwankungen, definierte er als jede Bewegung, die "nicht auf den Trend, die Saison oder die Konjunktur zurückgeht."⁶⁷ Wenn man diese Komponente als zufällig im Sinne von wandelbar und vielen einzelnen Faktoren ausgesetzt ansieht, so seien dies aber auch Eigenschaften, die auf die anderen Komponenten zutreffen könnten:

"Sobald wir also den Statistikern folgen und die Ursachen zufälliger Schwankungen erörtern, müssen wir offen zugeben, daß auch die säkularen, konjunkturellen und saisonmäßigen Schwankungen unregelmäßig verlaufen können, mit anderen Worten, will man säkulare, saisonmäßige und konjunkturelle Änderungen als regelmäßig behandeln, so gerät man vielfach in Widerspruch mit dem, was wir wissen, und noch mehr mit dem, was wir vermuten."⁶⁸

Dieser Ansatz wurde jedoch nicht konsequent zu Ende gedacht.⁶⁹ Zwar unterstellte Mitchell diese Kontingenz für alle Komponenten, glaubte aber andererseits, daß Trend-, Saison- und Konjunkturschwankungen durch mehrere Ursachen, die nach einem Gesetz wirken, in Erscheinung treten, für die Zufallskomponente dagegen kein Gesetz wirke – somit auch kein Verteilungsgesetz.

Korrelationen

In der bislang beschriebenen Entwicklung stand die univariate Analyse im Vordergrund. Andererseits fand natürlich auch die Beziehung zwischen mehreren (zwei) Variablen das Interesse der empirischen Ökonomen. 1922 widmeten W. F. Ogburn und D. S. Thomas eine Untersuchung der Auswirkung zyklischer Schwankungen des Wirtschaftslebens auf die Entwicklung sozialer Größen.⁷⁰ Sie setzten zyklische Schwankungen des ökonomischen Sektors voraus und fragten: "Do these fluctuations in business produce fluctuations in social conditions? Do we find relatively more births, deaths, marriages, and divorces in periods of business depression? Does crime and other social phenomena fluctuate with the business cycle?"⁷¹ Natürlich sei eine solche Gleichläufigkeit noch kein Beweis eines kausalen Einflusses; dennoch hätten vor allem die Untersuchungen von Mitchell und Persons gezeigt, daß mehrere ökonomische Reihen ihr Maximum erreichen, wenn der allgemeine Wohlstand hoch und einen Tiefpunkt durchlaufen, wenn der

⁶⁵ Ebda., S. 206.

⁶⁶ Vgl. ebda., S. 251-254.

⁶⁷ Ebda., S. 243.

⁶⁸ Ebda., S. 244.

⁶⁹ Ein derartiger variabler Komponentenansatz wurde erst sehr viel später methodologisch in einer stochastischen Konzeption umgesetzt.

⁷⁰ Vgl. Ogburn/Thomas (1922).

⁷¹ Ebda., S. 324.

allgemeine Wohlstand sehr niedrig sei. Um zumindest diese Gleichläufigkeit genauer untersuchen zu können, müsse zunächst der Trend eliminiert werden, wobei eine Parabel zweiter Ordnung sich als angemessen erwiesen habe. Ob dies der "wahre" Trend sei, könne letztendlich nicht entschieden werden: "As there was no mechanical test to show which particular line best described the true trend of the data, it was largely a matter of judgement of the eye."⁷² Im übrigen habe eine Korrelation der Residualreihen bei Ausschaltung des Trends durch einen neunjährigen gleitenden Mittelwert vergleichbare Resultate erbracht.

Auch hier stellte sich wieder die Frage, wie bereits bei der Diskussion zwischen Pearson, Yule, Hooker, Gosset und anderen,⁷³ welche Korrelation von Interesse war. Sollte der lang-, mittel- oder kurzfristige Zusammenhang untersucht werden? Die von Pearson und Gosset vorgeschlagene, in der Ökonomie von Persons empfohlene Vorgehensweise bestand darin, zuerst den Trend zu eliminieren und lediglich die Residuen zu korrelieren. Dies war sinnvoll, wenn die zugrundeliegenden Trends auf unterschiedliche Ursachen zurückzuführen waren. Wenn nun aber die Trends selbst die Beziehung definierten? Hier konnte man lediglich in Anlehnung an die theoretische Statistik konstatieren, daß das Problem eher außerhalb der Statistik gelöst werden sollte.⁷⁴

Unabhängig von der genauen Form der Trendbereinigung betrachteten die Autoren Korrelationen der relativen Abweichungen von diesen Trends, ohne die Residualreihe nochmals zu transformieren. Ihr Interesse war somit, wie schon von Yule und Hooker intendiert, eine mittelfristige Korrelation.

Ein wichtiger Aspekt wurde 1928 von M. Ezekiel angesprochen. Rückblickend auf die Entwicklung der Statistik stellte er fest, daß bezüglich der Profession der Statistik und der Ökonomie ein gegenseitiges Unbehagen bestehe. Während der Ökonom der Überzeugung sei, daß der Statistiker mit komplizierter Mathematik ein Material handle, das er nicht wirklich verstehe, mißtraue andererseits der Ökonom dem Statistiker, da er dessen Methoden nicht (mehr) verstehe.⁷⁵ Ein besonderes Problem sah er in der zu geringen Bedeutung der Dimension der Zeit in statistischen Untersuchungen. Die Zeit sei bislang nur insofern berücksichtigt worden, als man willkürliche zeitliche Schnitte vorgenommen habe, für die dann intern jeweils von einer Strukturkonstanz ausgegangen werde. Auch die Aggregationsebene spiele eine Rolle: Wenn etwa die täglichen Preise in Chicago und St. Paul betrachtet würden, so sei hier keinesfalls von einem einheitlichen "Markt" auszugehen; betrachte man dagegen monatliche Durchschnitte, würden die Differenzen bedeutend geringer, und wenn man schließlich jährliche Durchschnitte betrachte, so seien beide Märkte praktisch identisch. Eine exakte zeitliche und räumliche Beschränkung des Untersuchungsgegenstandes sei daher nötig.⁷⁶ Ins-

⁷² Ebd., S. 329.

⁷³ Siehe oben, Abschnitt A 4 b.

⁷⁴ Vgl. z. B. Roos (1936).

⁷⁵ Vgl. Ezekiel (1928, S. 200).

⁷⁶ "Exact delimitation of the 'market', both in time and space, thus has a very important bearing upon the significance of any statistical conclusions." Ezekiel (1928, S. 205).

besondere in langfristiger Sicht sei darüber hinaus von Änderungen im Beziehungsgefüge von ökonomischen Größen auszugehen, doch habe die Statistik bislang keine Methoden zur Untersuchung der “nature of relations between variables which are in shifting equilibrium with each other” zur Verfügung gestellt.⁷⁷ Dies sei jedoch eine zentrale Eigenschaft der meisten ökonomischen Phänomene:

“[...] it should be noted that the results obtained by statistical determination of the relations are not fundamental ‘laws of nature’ in the same sense as is the law of gravity. They are measures of the way that particular groups of men, in the aggregate have reacted to specific economic conditions during a specified period in the past. If the study is elaborate enough, it may even reveal the way in which the reaction has been changing during the period considered, and the direction and rate of change. But it does not tell how long the same reaction will continue to prevail, what new causes may arise to change the responses, or what the relations would be in the new situation.”⁷⁸

Eine statistische Methodologie, die dieser Tatsache nicht Rechnung trage, führe zu Resultaten, die unweigerlich Fehlurteile produzieren müßten.

In die gleiche Richtung wies die im selben Jahr an gleicher Stelle von S. Kuznets vorgetragene Kritik.⁷⁹ Kuznets Ansicht nach lag das Problem der Korrelation von Zeitreihen vor allem darin, daß die für die Berechnung notwendige Annahme konstanter struktureller Beziehungen während des betrachteten Zeitraums in der Regel nicht gerechtfertigt sei:

⁷⁷ Vgl. ebda., S. 218. Auf den dort angesprochenen konkreten Aspekt der Bestimmung von Angebots- und Nachfragekurven wollen wir hier nicht näher eingehen.

⁷⁸ Ebda., S. 223.

⁷⁹ Simon Smith Kuznets [1901-1985], geboren in Kharkov (Rußland), arbeitete zunächst bei einem statistischen Büro in der Ukraine und emigrierte 1922 in die USA, wo er an der *Columbia University* studierte (vor allem bei W. C. Mitchell) und 1926 in Ökonomie promovierte. 1927 wurde er Mitglied der Forschungsabteilung des *National Bureau of Economic Research* (NBER), wo er zunächst Mitchell bei seinen Konjunkturzyklenstudien assistierte. Zwar war er dort bis 1969 beschäftigt, seine Haupttätigkeit war jedoch seit den fünfziger Jahren Forschungen am *Social Science Research Council* (SSRC) gewidmet. Während dieser Zeit von 1930 bis 1954 war er Professor für Ökonomie und Statistik an der *University of Pennsylvania*, von 1942 bis 1944 *associate director* des *Bureau of Planning and Statistics of the War Production*, von 1954 bis 1960 Professor an der *Johns Hopkins University*, schließlich von 1960 bis zu seiner Emeritierung 1971 Professor an der *Harvard University*. 1949 war er Präsident der *American Statistical Association*, fünf Jahre später Präsident der *American Economic Association*. 1971 erhielt er den Nobelpreis für Wirtschaftswissenschaften für seine vergleichenden Analysen des wirtschaftlichen Wachstums. Eine Bibliographie seiner Publikationen findet sich in Kuznets (1989, S. 439-459). Kuznets, der “from 1930 onwards produced one damned, thick, square book after another on income and capital” (McCloskey (1987, S. 46)) hat die Entwicklung der makroökonomischen Wirtschaftsstatistik maßgeblich beeinflusst. Dieses Gebiet liegt jedoch außerhalb unseres Themas. Die hier behandelten frühen methodologischen Beiträge sind, abgesehen von einer russischen Publikation, die er im Alter von 20 Jahren verfaßte, seine ersten Veröffentlichungen überhaupt und haben auf sein späteres Hauptwerk nur zum Teil Auswirkungen gezeigt. In dem ersten, unmittelbar an die hier zu besprechenden methodologischen Beiträge anschließenden Werk (Kuznets (1930a)) entspricht die methodologische Argumentation zum großen Teil Kuznets (1928b). Als Trend wird hier in der Regel eine logistische Funktion unterstellt. Zu seinem späteren methodologischen Ansatz siehe allgemein Fogel (1989). Wir werden auf Kuznets noch einmal im Zusammenhang mit der Diskussion innerhalb der Wirtschaftsgeschichte zurückkommen.

"Whenever we correlate two time series we take them for a definite historical period. But are we sure that during the thirty-five or forty years covered by the data, the degree of concurrence has remained the same? Is it not more probable that most of the factors which made for a concurrence of the two phenomena have changed materially during so long a period? And if this be true, then the coefficient of correlation computed for this period is but an average of a changing trend in covariance, an average of little descriptive and still smaller inferential value."⁸⁰

Kuznets schlug daher vor, den untersuchten Zeitabschnitt in zwei Teile zu zerlegen und einen Korrelationskoeffizienten für jeden Unterabschnitt zu berechnen. Falls die Beziehung zweier Variablen zeitlich nicht konstant sein sollte, müßte sich dann ein Unterschied zeigen. Allerdings sei diese Methode zu grob, wenn man daran interessiert sei, welchen Pfad die Beziehung im Laufe der Zeit eingeschlagen habe. In diesem Fall sei ein "gleitender" Korrelationskoeffizient zu verwenden. Verschiedene Möglichkeiten bestünden hier: Am vielversprechendsten sei es, analog zu einem gleitenden Mittelwert die Korrelation jeweils für ein Intervall zu berechnen, z. B. 25 Jahre, anschließend ein Jahr hinzuzunehmen und den ersten Wert wegfällen zu lassen usw. bis das Ende der Reihe erreicht sei. Der Wert des Korrelationskoeffizienten für ein Intervall könne dann in einer Abbildung jeweils in der Mitte des Intervalls abgetragen werden. Diese Vorgehensweise bedeute keine grundsätzliche Abkehr von der traditionellen Vorgehensweise: "It does not introduce any changes into the mathematics or logic of the Pearsonian r ."⁸¹ Da Kuznets die Korrelationen von Zyklen interessierten, mußten die Reihen (Produktions- und Preisreihen verschiedener Produkte und Länder) zunächst anhand von Funktionen trendbereinigt werden. Die auf die Restkomponente angewandte gleitende Korrelation erbrachte für die Beziehung der Roheisenproduktion und -preise in fünf verschiedenen Ländern die in folgender Abbildung wiedergegebenen "gleitenden" Koeffizienten.⁸²

⁸⁰ Kuznets (1928a, S. 121).

⁸¹ Ebda., S. 124.

⁸² Kuznets wies hier zu Recht darauf hin, daß die Wahl der Länge der jeweiligen Abschnitte nicht statistisch bestimmt werden könne: "On the one hand, we do not want to make the period too long, for fear that it might cover periods through which the economic conditions determining covariance changed considerably. On the other hand, we need a considerable number of items, a number sufficient to prevent any vigorous accidental disturbances that might obscure the underlying secular movement in the degree of correlation. The difficulty is obviously of the kind that does not allow for solution by any set up criterion, but requires an experimental procedure of trial and error." Ebda., S. 125.

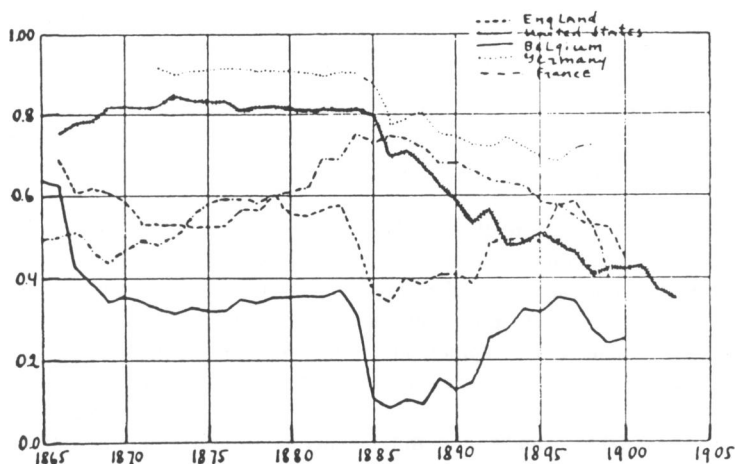


Abb. 7: Gleitende Korrelation zwischen Roheisenproduktion und Preisen in fünf Ländern 1865 bis 1903. Quelle: Kuznets (1928a, S. 134).

Diese und die noch für mehrere andere Reihen vorgenommenen Berechnungen ließen Kuznets resümieren: Lediglich eine der annähernd dreißig von ihm untersuchten Beziehungen wies eine in der Zeit stabile Strukturkonstanz auf. Insofern sei die Annahme struktureller Stabilität zwischen ökonomischen Größen im Verlauf der Zeit eher die Ausnahme. Sein Fazit lautete:

“Thus, before statistical analysis of time series can be used as a statistical complement of static economic theory, extreme care should be taken to eliminate all the historical or dynamic elements from the analyzed time sequence. [...] a straight correlation of series does not eliminate the historical change of factors that determine covariance. The empirical element with which quantitative theory has to deal is more historical than it would seem at first glance. And there is a corresponding need for a material change of the somewhat too static methods of a biometric frequency statistical mechanism.”⁸³

c. Die Rolle der Zufallsgrößen

Während Mitchell 1927 eine deskriptive, die Einmaligkeit jedes einzelnen Zyklus betonende Vorgehensweise forderte, erschienen im selben Jahr zwei Arbeiten, die einen fundamentalen Wandel der Sichtweise von Zyklen bewirken sollten: dies waren die in Abschnitt A 4 b bereits behandelten Arbeiten von Yule und Slutsky. Unmittelbar nach dem Erscheinen von Slutskys Arbeit erkannte S. Kuznets als erster die Tragweite des Beitrags. C. J. Bullock, W. M. Persons und W. L. Crum hatten in einer Antwort auf eine neue Interpretation des Harvard Business Index durch K. Karsten darauf aufmerksam gemacht, daß durch die Summation einzelner Elemente eine wellenförmige Erscheinung auftreten könne, die dem Aussehen

⁸³ Ebda., S. 136. Hervorhebung von uns.

ökonomischer Zyklen entspräche. Kuznets ging in Anlehnung an Slutsky einen Schritt weiter: "It is not only 'possible' that a summation of a random series will yield cycles, but also quite certain that this will be the case."⁸⁴ Diese Einsicht sei eben jenem bemerkenswerten Beitrag Slutskys zu verdanken, der nachgewiesen habe, daß 1. zyklische Prozesse durch Summierung voneinander unabhängiger Zufallszahlen entstehen könnten und 2. diese Zyklen eine gewisse Regularität aufweisen.

Kuznets hielt insbesondere die erste These für bedeutend und zeigte im folgenden die Konsequenzen auf, die seiner Ansicht nach damit für die ökonomische Forschung verbunden waren. Zunächst verwies er auf das erste von Slutsky angeführte Beispiel der 10gliedrigen Summe, aus der "a line of well-marked cyclical swings" resultiere, die eine solch frappierende Ähnlichkeit mit dem *Quarterly Index of Business Conditions* von 1855 bis 1877 aufweise (Abb. 8).⁸⁵ Zur Illustration führte er die Experimente von Slutsky selbst durch. Zum einen mit gleichverteilten (analog zu Slutsky), zum anderen mit annähernd normalverteilten Zufallszahlen: der Größe "of the meal check for a noonday meal in the Common Cafeteria, Columbia University".

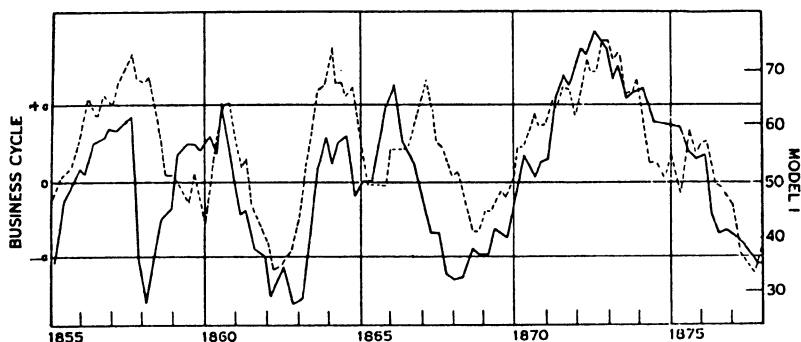


Abb 8: "Quarterly Index of Business Conditions" 1855 bis 1877 und aufsummierte Zufallsgrößen. Quelle: Slutsky (1937, S. 110).

Im Gegensatz zu Slutsky begründete Kuznets dieses Phänomen nun aber nicht mit dessen Argumenten, sondern mit einigen – recht kuriosen – kombinatorischen Überlegungen über Vorzeichenwechsel.⁸⁶ Er gelangte zu den folgenden qualitati-

⁸⁴ Kuznets (1929, S. 285). Er lehnte eine vollständige Summation auch ab, da sie unrealistisch sei: "We shall omit the straight cumulation altogether, because it is too far removed from reality." (S. 269).

⁸⁵ Ebda., S. 286. Slutskys Abbildung ist an dieser Stelle von Kuznets reproduziert. Er betonte den Unterschied, der zu der Vorgehensweise von Bullock, Persons und Crum besteht, da hier nicht alle Terme kumuliert wurden.

⁸⁶ Vgl. ebda., S. 263-266. Soweit wir sehen, findet diese Interpretation später auch keine Rezeption mehr. Daß ihm die Zusammenhänge nicht völlig klar waren, zeigt eine Aussage

ven Schlußfolgerungen: 1. Summen von Zufallsvariablen führten zu Zyklusbildungen, die ökonomischen Daten ähneln, 2. diese seien auf Ketten gleichen Vorzeichens oder Ausreißer zurückzuführen, 3. Amplitude und Periode der Zyklen hingen von der Schiefe der Verteilung der Zufallsvariable und der Anzahl der aufsummierten Zufallsgrößen ab und 4. eine Gewichtung "with its central items weighted more heavily than the extreme ones" (S. 273) impliziere einen glatteren Verlauf als eine einfache Gewichtung. Zur Beurteilung dieser Ergebnisse müsse man zunächst fragen: Wenn die Summierung zufälliger Ereignisse zyklische Verläufe erzeugt, ist dann auch umgekehrt der zyklische Verlauf ökonomischer Reihen auf aufsummierte Zufallsereignisse zurückzuführen? Kuznets Ansicht nach ist ein solcher Umkehrschluß nur dann zulässig, wenn man dies im Zusammenhang mit anderen Hypothesen abwägend betrachtet (S. 274). Er gelangte daher zu dem Schluß, daß bei statistischen Analysen der seriellen Korrelation Rechnung getragen werden müsse und man daran denken könne, die Art der Summation in einer vorliegenden Reihe zu identifizieren, um so den Erzeugungsmechanismus zu bestimmen.⁸⁷ Kuznets sprach aber zugleich eine Warnung aus:

"However, the economic conditions that underlie time series suppose that a chance distribution of events occurs only within narrow limits and does not permit us to treat the rises and declines in a time series as continuous cumulations of economic events that are continuously random in their character."⁸⁸

Falls der Umkehrschluß aber trotz dieser Bedenken zutreffen sollte, so hätte dies weitreichende Konsequenzen für die bisherigen Konjunkturtheorien:

"If cycles arise from random events, assuming the summation of the latter, then we obviously do not need the hypothesis of an independent regularly recurring cause which is deemed necessary by some theorists of business cycles. Indeed if one can explain how in certain processes of economic life, the response to stimuli is cumulative, then the whole discussion of the cause of business cycles becomes supererogation."⁸⁹

Dieser Tatsache sollten zukünftige Erklärungen Rechnung tragen.⁹⁰

hinsichtlich seines ersten Beispiels, in dem er zunächst einen fünfgliedrigen "moving average" verwendete, sich anschließend aber für die Präsentation eines elfgliedrigen Mittelwertes entschloß, "[...] because the five-item average did not indicate the resulting cycles as clearly as desired" (S. 262). Außerdem ist nicht festzustellen, ob er tatsächlich wie Slutsky die Summe zurückliegender Werte benutzt hat oder den häufig verwendeten zentrierten "gleitenden Mittelwert".

⁸⁷ Dies ist nichts anderes als das Vorgehen der modernen ARIMA-Zeitreihenanalyse.

⁸⁸ Kuznets (1929, S. 274).

⁸⁹ Ebda.

⁹⁰ Aus Sicht der gegenwärtigen Diskussion um die sogenannten unit-root-Tests (siehe dazu Metz (1995)) ist dies geradezu als visionär zu bezeichnen. Seine zweite Folgerung betrifft die "Glattheit" der zyklischen Erscheinungen, die etwa beim Großhandelspreisindex zu beobachten und ebenfalls auf die wiederholte Durchschnittsbildung zurückzuführen sei. Hier ging er jedoch irrtümlich davon aus, daß eine *Querschnittsdurchschnittsbildung* (von mehreren Produkten zu jeweils einem Zeitpunkt) mit einer Summation von zurückliegenden Zeitpunkten übereinstimmt, was zweifellos nicht der Fall ist.

Unmittelbar darauffolgende Arbeiten sind mit dem Namen R. Frischs verbunden.⁹¹ Vor seinen berühmten "ökonometrischen" Studien von 1933 und 1934 widmete Frisch sich der univariaten Zeitreihenanalyse, wobei er sich vor allem dem Aspekt der Komponentenzerlegung zuwendete.⁹² Der Ausgangspunkt für seine Beschäftigung mit dieser Thematik war, daß seiner Ansicht nach keine Methodologie einer Komponentenanalyse existierte, sondern statt dessen eine Reihe logisch unverbundener Verfahren.⁹³ Es sei zwar angemessen, eine Reihe als Summe zyklischer Komponenten anzusehen,⁹⁴ doch sei eine herkömmliche Periodogrammanalyse nicht anwendbar, da ökonomische Reihen sich durch *variable Zyklenlängen* auszeichneten, die Periodogrammanalyse aber konstante Zyklen voraussetze.⁹⁵ Statt dessen schlug er eine lokale Analyse vor, die nichtparametrische Anpassungen zu einem Zeitpunkt t aufgrund von Informationen in der Nähe dieses Punktes liefern sollte. Unabhängig von der konkreten Ausgestaltung des Verfahrens⁹⁶ ist in vorliegendem Zusammenhang vor allem von Interesse, daß Frisch hier 1. von zeitvarianten Strukturen ausging und 2. einem stochastischen Charakter vorliegender Zeitreihen explizit Rechnung tragen wollte. Insbesondere die schon in der russischen Fassung⁹⁷ unter empirisch orientierten Ökonomen wie Frisch große Unruhe auslösende Arbeit von Slutsky motivierte ihn, nach einem allgemeinen Testverfahren zur Unterscheidung von "künstlichen" und "echten" Zyklen zu suchen. Die zwei Jahre später erschienene Arbeit *Propagation problems and impulse problems in dynamic economics* ging konsequenter in diese Richtung und machte Frisch international bekannt.⁹⁸ Frisch untersuchte hier, welche Eigenschaften ein

⁹¹ Ragnar Anton Kittil Frisch [1895-1973] promovierte in mathematischer Statistik und brachte den größten Teil seines akademischen Lebens an der Universität Oslo. Er war eines der Gründungsmitglieder der *Econometric Society* (s. u.). Vgl. Morgan (1990, S. 83f) mit weiterführender Literatur. Er erhielt 1969 zusammen mit J. Tinbergen den ersten Nobelpreis für Wirtschaftswissenschaften. Frischs Arbeiten sind in der Regel schwer zu lesen, da sie in einer sehr technischen Sprache abgefaßt sind und eingeführte Begriffe oftmals in einer anderen Bedeutung verwenden. Eine Reihe seiner Schriften wurde jetzt von Bjerkholt (1995) herausgegeben.

⁹² Vgl. Frisch (1931), zum folgenden auch Morgan (1990, S. 83ff). Die Studie Frischs war Teil eines umfassenden, mimeographierten Manuskriptes (von immerhin 248 Seiten Umfang, vgl. Wold (1965a, S. 147)) mit dem Titel "Time series analysis", das auf an der Universität Oslo gehaltenen Vorlesungen beruhte. Dieses Manuskript wurde von Schumpeter (1939 [1961, S. 218ff]) für seine Darstellung der zeitreihenanalytischen Aspekte zugrundegelegt.

⁹³ Vgl. dazu auch Boumans (1995, S. 132).

⁹⁴ Trends sollten als Teil eines längeren Zyklus angesehen werden, erratische Schwankungen als kurzfristige Zyklen, deren Zyklenlänge kürzer als die zugrundeliegende Zeiteinheit ist.

⁹⁵ Zwar hatte Yule (1927) etwa zur gleichen Zeit mit seiner grundlegenden Arbeit über AR(2)-Prozesse die Möglichkeit zur Modellierung solcher "stochastischer" Zyklen gegeben, doch ihre Spektraldarstellung war, soweit wir sehen, erst durch die Arbeit von Wold (1938) eingeführt worden.

⁹⁶ Vgl. dazu anschaulich Morgan (1990, S. 85ff).

⁹⁷ Die ursprüngliche Version von Slutsky (1937) erschien bereits 1927 in russischer Sprache mit einem englischsprachigen Summary.

⁹⁸ Samuelson ist der Ansicht, daß die Bedeutung dieser Arbeit, die einen konzeptionellen Wandel von statischen zu dynamischen stochastischen Modellen bewirkte, mit dem Übergang von der klassischen Physik zur Quantenmechanik verglichen werden kann. Vgl. Wallis (1994b, S. x).

dynamisches System aufweisen müsse, damit es solche Zyklen erzeuge, die in der Realität zu beobachten seien.⁹⁹ Bei der Wahl bestimmter Koeffizienten erzeugt sein Modell deterministische Zyklen mit einer gedämpften Amplitude, die durch die Einwirkung externer Einflüsse zu permanenten "stochastischen" Zyklen führen.¹⁰⁰ Die Größe der Koeffizienten ermittelte er nicht durch eine statistische Schätzung, sondern durch eine "Abwägung": als numerische Größen, die in etwa denjenigen Größen entsprechen sollten, die man in der gegenwärtigen ökonomischen Realität erwarten würde:

"At present I am only guessing very roughly at these parameters, but I believe that it will be possible by appropriate statistical methods to obtain more exact information about them. I think, indeed, that the statistical determination of such structural parameters will be one of the main objectives of the economic cycle analysis of the future. If we ask for a real *explanation* of the movements, this type of work seems to be the indispensable complement needed in order to co-ordinate and give a significant interpretation to the huge mass of empirical descriptive facts that have been accumulated in cycle analysis for the past ten or twenty years."¹⁰¹

Entscheidend sei bei diesem Modell, daß die Länge der Zyklen bei einer bestimmten Wahl der Parameter sehr genau mit den von der empirischen Konjunkturforschung festgestellten Zyklenlängen übereinstimme.¹⁰²

Einen vorläufig letzten Schritt in der Einführung einer aktiven Rolle von Zufallsgrößen in die Ökonomie bildete schließlich die Arbeit von H. Working. Working berichtete 1934, daß bei einer Reihe von Rohstoff- und Aktienpreisreihen "zufällige" *Differenzen* beobachtet wurden. In solchen Fällen sei die übliche Komponentenzerlegung nicht angemessen:

"These concepts are inappropriate and misleading when applied to cases in which the dominant tendency is for the effects of successive events to be independent and cumulative. An outstanding characteristic of this type is that its changes are largely random and unpredictable. [...] Brief study of the charts which follow will show that in a series composed by purely random changes (a strict random-difference series) conspicuous trends will be found. Such 'trends', however, must be regarded merely as generalized descriptions of the course of the series over a certain period, not as norms, nor as bases for predicting the future course of the series even the briefest subsequent period. In such a series also, frequent movements of indubitable cyclical confirmation occur, but they cannot be regarded as evidence of true cycles – even irregular cycles – unless that term be divorced from its usual connotation of movement that tends to generate, or at least to reflect influences that tend to generate, reverse movement."¹⁰³

⁹⁹ Das Modell bestand aus drei gemischten Differential- und Differenzgleichungen. Vgl. dazu auch ebda.

¹⁰⁰ Dies ist das von Yule (1927) eingeführte Prinzip. Zu den technischen Details vgl. Rahlf (1996b).

¹⁰¹ Frisch (1933a, S. 185). Siehe auch Morgan (1990, S. 94). Diese Vorgehensweise erhält in letzter Zeit in der Ökonometrie erneute Aufmerksamkeit unter der Bezeichnung "Calibration". Vgl. Hendry/Morgan (1995b, S. 49).

¹⁰² Die Amplitude der Zyklen hing hierbei nicht von den Parameterwerten, sondern den Ausprägungen der Zufallsgrößen ab. Siehe dazu Rahlf (1996b).

¹⁰³ Working (1934, S. 12).

Wie bereits Yule und Slutsky, auf dessen Vorarbeit er sich berief, verwendete er "Simulationen" als Vergleichsmaßstab für real existierende Reihen. Zufallsziehungen und sukzessive Aufsummierungen der Zahlen ergaben die in Abb. 9 wiedergegebenen künstlichen Reihen.

Wie man unmittelbar erkennen könne, bestehe hier die Gefahr, in derartigen Reihen irrtümlich Zyklen und überlagerte Zufallsschwankungen zu sehen. Diese seien jedoch nicht "real", sondern nur durch die Aufsummierung reiner Zufallszahlen entstanden, und somit gäbe es weder einen Trend noch Zyklen im oben definierten Sinne. Vergleiche man diese Reihen mit der Weizenpreisreihe, so sei eine ähnliche Struktur durchaus zu erkennen, wobei letztere eher aufsummierte Zufallsgrößen aus einer schiefen Verteilung sowie mit einer vermutlich systematisch variierenden Varianz darstellten.¹⁰⁴

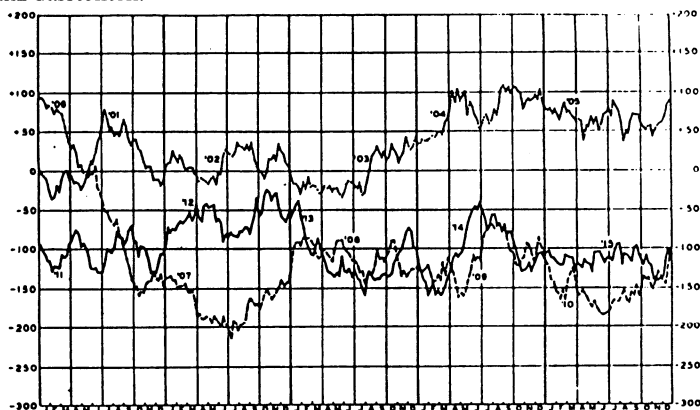


Abb. 9: Beispiele für aufsummierte Zufallsreihen. Quelle: Working (1934, S. 20).

Die Rezeption der Arbeiten von Yule und Slutsky durch Frisch und Working war der Beginn eines fundamentalen Richtungswechsels in der statistischen Analyse von ökonomischen Zeitreihen, der zunächst eine völlige Abkehr von den traditionellen Komponentenmodellen bedeutete. Wie sehr diese andere Rolle der Zufallsgrößen noch mehrere Jahre nach ihrer Einführung selbst unter den Gründungsvätern der Ökonometrie mißverstanden wurde, belegt eine Notiz von G. Tintner aus dem Jahr 1938.¹⁰⁵ Tintner, der sich im übrigen bei T. W. Schultz, G. Stigler, H.

¹⁰⁴ Zum Vergleich einer Weizenpreisreihe mit aufsummierten Zufallszahlen untersuchte er die Verteilung von maximalen, fünf Zeiteinheiten auseinanderliegenden Differenzen beider Reihen. Das Varianzproblem wurde erst sehr viel später (im Zusammenhang mit trendfreien Reihen) wieder aufgegriffen. Siehe dazu den Überblick von Thome (1994c).

¹⁰⁵ Gerhard Tintner [1907-1983] verbrachte nach seiner Promotion 1929 in Wien bis 1936 Studienjahre in Harvard, Columbia, Chicago, Stanford, Berkeley und Cambridge, anschließend kurze Zeit am Institut für Konjunkturforschung in Wien, ging 1937 an die *Iowa State University*, wo er schließlich Professor für Ökonomie, Mathematik und Statistik wurde. 1962 wurde er Professor in Pittsburgh, 1963 bis 1973 war er Professor an der *University of Southern California*, anschließend Professor und Direktor des Instituts für Ökonometrie der Technischen Universität Wien. Er war zeitweise Mitherausgeber der *Econometrica*, zwi-

Hotelling und C. F. Roos für ihre Hilfe und Unterstützung bedankte,¹⁰⁶ versuchte hier, eine ökonomische, auf Störungen in Erwartungen und rationalem Verhalten zurückgehende Begründung für die Theorie der Fehler in Zeitreihen zu entwickeln.¹⁰⁷ Tintner unterschied zwei Arten von Fehlern, die auf unterschiedliche Ursachenkonstellationen zurückgeführt werden konnten, wobei der erste Fehlertyp mit demjenigen übereinstimme, der nach den üblichen Komponentenzersetzungen als Restkomponente verbleibe, nachdem der kausale Teil der Reihe eliminiert worden sei.¹⁰⁸ Der zweite Fehlertyp, der fundamentalere sei und einen tieferen Einfluß auf das ökonomische Geschehen ausübe (beide Fehlertypen schlossen sich im übrigen nicht aus, sondern ergänzten sich), sei eher mit den von R. Frisch und anderen postulierten Zufallsfluktuationen, die einen Startmechanismus für Zyklen darstellten, zu vergleichen. Nach dieser Theorie befinde sich ein ökonomisches System in einem instabilen Gleichgewicht, in dem selbst kleine Störungen große zyklische Schwankungen hervorrufen könnten.

Davon zu unterscheiden sei eine *andere* interessante Theorie, die der russische Statistiker Slutsky entwickelt habe. Demnach könnten Konjunkturschwankungen durch eine spezielle Aufsummierung von Zufallsgrößen hervorgerufen werden, wie ein Vergleich derart erzeugter Reihen mit realen konjunkturellen Wirtschaftsreihen zeige. Aber:

“His theory lacks, however, any explanation of the mechanism through which small random variations are ‘summed’ in the economic process. It cannot be denied that random variations occur and that there is a striking similarity between the appearance of economic time series and the summation series of a random series. But until it is shown just how random variations influence the economic system, we must regard the similarity as merely spurious and superficial and class this theory with the sun-spot theory as an explanation of business cycle phenomena. Correlations and apparent similarities, however striking, cannot be considered convincing until further proof is offered. This statement holds especially with regard to time series. Mere correlations without sufficient theoretical foundations fail to carry conviction.”¹⁰⁹

Frisch sah sich genötigt, in einer Antwort dieses Mißverständnis aufzuklären. Sein Modell sei stabil, da die Dynamik ohne Zufallseinflüsse einer *gedämpften* Sinusschwingung entspreche.¹¹⁰ Diese Dynamik werde durch den Einfluß der Zufallsgrößen nicht instabil, da die “Schocks” durch den Mechanismus (d. h. durch entsprechende Größen der Koeffizienten) absorbiert werden. Entscheidend sei nun der Umstand, den er in besagtem Aufsatz auch mehrfach betont habe: “*This idea*

schen 1933 und 1952 einer ihrer häufigsten Autoren und verfaßte unter anderem mehrere Lehrbücher zur Zeitreihenanalyse und Ökonometrie.

¹⁰⁶ Charles F. Roos und Harold Hotelling (der auch grundlegende Beiträge zur theoretischen Statistik publiziert hat) waren gelehrte Mathematiker.

¹⁰⁷ Er sprach damit auch ein Problem an, das zu dieser Zeit des öfteren diskutiert wurde: inwieweit etwa Prognosen das Verhalten von Wirtschaftssubjekten beeinflussen und damit wiederum die Prognosen korrigiert werden müssen. Dieser Punkt wurde später in der berühmten Lucas-Kritik wieder aufgegriffen.

¹⁰⁸ Vgl. Tintner (1938, S. 147).

¹⁰⁹ Ebda., S. 148.

¹¹⁰ Siehe dazu Rahlf (1996b).

is fundamentally the same as Slutsky's and Yule's."¹¹¹ Es sei zwar richtig, daß Slutsky den von ihm beschriebenen Mechanismus nicht in den Rahmen einer ökonomischen Theorie eingebettet habe und nicht erklärte, aus welchem Grund die Schocks sich aufsummieren sollten. Genau dieser Punkt sei aber durch sein eigenes *Propagation*-Paper geklärt worden: Wenn eine Gruppe von Variablen durch ein lineares dynamisches System definiert und dieses System durch Schocks gestört werde, dann ergäbe sich zwangsläufig für die Schocks ein Verlaufsmuster, "which would have given the time evolution of this variable, if no shocks had occurred. Thus economic theory furnishes the weight system, statistical theory does the rest."¹¹²

d. Der traditionelle Weg

Wenn bereits mathematisch gebildete Ökonomen und Vorreiter einer Mathematisierung und Quantifizierung der Ökonomie Verständnisprobleme bezüglich der im letzten Abschnitt geschilderten Konzepte hatten, so verwundert es nicht, wenn sich in der Folgezeit eine Tradition entwickelte, die andere Wege ging. Die ein Jahr nach Tintners "Note" erschienene und in methodischer Hinsicht traditionelle, monumentale Arbeit von Joseph A. Schumpeter¹¹³, *Business Cycles*, mit dem bezeichnenden Untertitel *A Theoretical, Historical and Statistical Analysis of the Capitalist Process*, markierte bereits eine Richtung, die die im vorigen Abschnitt beschriebene Entwicklung nicht mehr ernsthaft rezipierte.¹¹⁴

Schumpeter verstand diese Arbeit als eine Fortsetzung bzw. Einlösung seiner bereits 1910 erschienenen Theorie der wirtschaftlichen Entwicklung, die er nun mit historischem und statistischem Material zu einer umfassenden Analyse ergänzen wollte.¹¹⁵ Er betonte, daß sich zwar eine Reihe von Phänomenen der Wirtschaftslage ohne weitere konzeptionelle Überlegungen erfassen ließe, eine weitergehende Analyse aber darüber hinausgehen müsse und vor allem historische Untersuchungen notwendig seien:

¹¹¹ Frisch (1939, S. 639). Hervorhebung von uns.

¹¹² Ebda., S. 640. Der Ausdruck "shock" wurde im übrigen in Analogie zur Physik gewählt, wo solche Modelle verbreitet waren.

¹¹³ Joseph Alois Schumpeter [1883-1950] studierte seit 1901 an der Juridischen Fakultät Wien und promovierte dort 1906 zum Doktor beider Rechte, ging anschließend nach Oxford und Cambridge und arbeitete 1907/1908 als Anwalt am Internationalen Gerichtshof in Kairo. In dieser Zeit veröffentlichte er sein erstes größeres Werk (im Alter von 25 Jahren), *Das Wesen und der Hauptinhalt der theoretischen Nationalökonomie*, das er auch als Habilitationsschrift einreichte. 1909 wurde er Professor an der Universität Czernowitz, 1911 an der Universität Graz. 1919 wurde er Staatssekretär für Finanzen (Finanzminister), 1921 Präsident der Biedermann-Bank (die drei Jahre später zusammenbrach), 1925 Professor für Finanzwissenschaft in Bonn, 1932 bis 1950 Professor in Harvard, nachdem eine Bewerbung in Berlin gescheitert war. 1937 bis 1941 war Schumpeter Präsident der *Econometric Society*, 1948 Präsident der *American Economic Association*, 1949 der erste Präsident der *International Economic Association*. Sein wissenschaftliches Leben war von einer intellektuellen Rivalität mit J. M. Keynes gekennzeichnet.

¹¹⁴ Wir verwenden im folgenden die deutsche Übersetzung, Schumpeter (1939 [1961]).

¹¹⁵ Wir wollen uns in diesem Zusammenhang auf den methodologischen Aspekt beschränken.

“Insbesondere sollte es einleuchten, daß wir so weit wie möglich in die Vergangenheit zurückgehen müssen – weil wir kein anderes Mittel haben, um eine größere Anzahl von einzelnen Schwankungsverläufen zu beobachten – und daß daher die historische Forschung von allerhöchster Bedeutung ist, selbst wenn wir es mit den praktischen Problemen der Gegenwart zu tun haben.”¹¹⁶

Schumpeter sprach sich dabei sowohl gegen die bislang praktizierten induktiven als auch deduktiven Vorgehensweisen aus.¹¹⁷ Eine statistische Untersuchung allein könne nicht zum Ziel führen, denn

“[...] kein statistischer Befund kann je einen Lehrsatz beweisen oder widerlegen, an den wir kraft einfacherer und grundsätzlicherer Tatsachen begründeterweise glauben. Er kann einen solchen Satz nicht beweisen, weil sich ein und dasselbe Verhalten einer Zeitreihe analytisch auf unendlich viele Weisen erklären läßt. Er kann den Satz nicht widerlegen, weil eine höchst reale Beziehung so stark von anderen auf das untersuchte statistische Material wirksamen Einflüssen überlagert sein kann, daß sie im Zahlenbild völlig verschwindet oder aber dadurch an Bedeutung für unser Verständnis des Falles einzubüßen droht. Daraus folgt, daß der Anspruch, der üblicherweise für die statistische Induktion und Beweisführung erhoben wird, präzisiert werden muß. Ein Material, das wie das unsere so vielen Störungen unterliegt, erfüllt nicht die logischen Erfordernisse des Induktionsverfahrens.”¹¹⁸

Letztendlich bestehe ein unauflösbarer Gegensatz zwischen historischen Ereignissen, die per definitionem einmalig sind und analytischen, generellen und theoretischen Aussagen. Insofern müsse es die erste Aufgabe des Forschers sein, die in Diagrammen ökonomischer Zeitreihen erkennbare “unregelmäßige Regelmäßigkeit” zu beschreiben. Methodologisch finden sich bezüglich dieser Aufgabenstellung jedoch keine weiteren Ausführungen. Die Analyse beschränkte sich diesbezüglich auf graphische Darstellungen von Trendfunktionen oder gleitenden Mittelwerten.¹¹⁹

W. C. Mitchell setzte dagegen seinen Ansatz der statistischen Messung konkreter Zyklen systematisch fort. 1946 kulminierte diese Arbeit im zweiten Band der Konjunkturzyklenstudien des NBER,¹²⁰ der monumentalen, zusammen mit A. F. Burns¹²¹ verfaßten Monographie *Measuring Business Cycles*.¹²² Darin wurden von

¹¹⁶ Schumpeter (1939 [1961, S. 36]).

¹¹⁷ Vgl. ebda., S. 37-39.

¹¹⁸ Ebda., S. 39. Er sah sich mit dieser Argumentation in einer Linie mit W. C. Mitchell, W. Persons und A. Spiethoff.

¹¹⁹ Schumpeter ging zwar auf die Periodogrammanalysen ein (S. 175), betonte aber wie Mitchell, daß die Ergebnisse nicht überbewertet werden sollten. Bezeichnenderweise unterlag er bezüglich der Arbeit von Frisch (1933a) demselben Mißverständnis wie bereits Tintner (1938). Er kommentierte Frischs Ansatz: “Obwohl er die Arbeiten Wicksells und Slutzkys als Ausgangspunkt anführt, ist sein Argument in Wirklichkeit ein ganz anderes” (S. 191), wobei er auf den von Frisch in dessen Arbeit von 1931 verwendeten Begriff der “wechselnden Harmonik” verwies. Der weitgehende Verzicht auf weiterführende statistische Verfahren war ein Hauptkritikpunkt von Kuznets Besprechung der Arbeit, vgl. Kuznets (1940).

¹²⁰ Der erste Band war Mitchell (1927).

¹²¹ Arthur Frank Burns [1904-] war zunächst der Nachfolger von S. Kuznets als Assistent von W. C. Mitchell. Später lehrte er an den Universitäten von Columbia und Rutgers. Er diente den Präsidenten Eisenhower (1953-1956) und Nixon (1969-1970) als Wirtschaftsberater, war von 1970 bis 1978 *Chairman* des *Board of Governors of the Federal Reserve System*

den Autoren nicht weniger als 1277 Reihen der englischen, französischen, deutschen und US-amerikanischen Wirtschaftsentwicklung seit Ende des letzten Jahrhunderts vollständig nach einem eigens entwickelten Schema untersucht,¹²³ über 80 % davon auf Monats- oder Quartalsbasis. Zusätzlich wurde eine Reihe weiterer Serien herangezogen, die nur für Teilaspekte ausgewertet wurden. Wir wollen die zugrundeliegende Methodologie hier etwas ausführlicher darstellen. Sie wurde immerhin für Jahre hinaus maßgeblich für die "offizielle" Konjunkturforschung der USA.

Burns und Mitchell unterschieden "spezifische Zyklen" und "Referenzzyklen".¹²⁴ Zunächst wurden für jedes der vier untersuchten Länder Referenzdaten anhand einer vormals von W. L. Thorp für das NBER zusammengestellten Kompilation, die von Burns und Mitchell "refined, tested, and at need amended" wurde, festgestellt.¹²⁵ Diese Daten definieren, jeweils von einem Tiefpunkt zum nächsten, einen "Referenzzyklus". Anschließend wurden für die – saisonbereinigte – Reihe der Mittelwert innerhalb jedes Referenzzyklus berechnet und die Daten in Beziehung zu diesem gesetzt; diese Werte wurden von ihnen "reference-cycle relatives" genannt.¹²⁶

Der nächste Schritt bestand darin, für jede einzelne Serie spezifische Zyklen festzustellen. Auch hier wurden wiederum die Durchschnittswerte während der einzelnen Zyklen berechnet und die Werte der Serie relativ zu diesen Durchschnitten umgerechnet.¹²⁷ Aus diesen Werten wurden anschließend zwei Arten von zusammenfassenden Tabellen erstellt: fünf sogenannte "S-Tabellen" und drei bzw. vier "R-Tabellen". Der erste Tabellentyp (S1) notiert für jeden Start- und Endzeitpunkt eines spezifischen Zyklus einer Reihe (1)¹²⁸ die Anzahl der Monate, die der jeweilige Hochpunkt dem entsprechenden Zeitpunkt des Referenzzyklus hinterherhinkt oder vorweilt (2), das Datum des Hochpunktes des Referenzzyklus (3), das gleiche für den Tiefpunkt (4, 5), die Anzahl der Monate (die Dauer) der Expansionsphase (6) und Kontraktionsphase (7) und des gesamten spezifischen Zyklus (8 = 6+7), die entsprechenden Vor- oder Nachläufe in Monaten gegenüber dem Referenzzyklus (9, 10, 11), schließlich den prozentualen Anteil der Dauer der

und von 1981 bis 1988 Botschafter der USA in der Bundesrepublik Deutschland. Zu Mitchell siehe oben, Anm. 19.

¹²² Burns/Mitchell (1946).

¹²³ Allerdings ist sowohl der Beginn als auch das Ende der einzelnen Reihen sehr unterschiedlich. Siehe die Übersichten in Burns/Mitchell (1946, S. 20).

¹²⁴ Vgl. ebda., S. 24ff.

¹²⁵ Ebda., S. 24. Die genaue Vorgehensweise läßt sich nicht rekonstruieren. Nach Angabe der Autoren ist die Datierung für die USA sicherer als für die übrigen Länder, außerdem für spätere Jahre sicherer als für weiter zurückliegende. Siehe dazu die Ausführungen in Mitchell (1927, passim, v. a. Kap. IV).

¹²⁶ Der Grund für diese Vorgehensweise ist eine dadurch ihrer Ansicht nach herbeigeführte Vergleichbarkeit aller Reihen und Werte. Vgl. Burns/Mitchell (1946, S. 24).

¹²⁷ Das bedeutet, daß aus jeder einzelnen der über tausend Reihen zwei Relativreihen berechnet wurden.

¹²⁸ In Klammern jeweils die Spalte.

Expansionsphase und Kontraktionsphase an der Gesamtdauer des spezifischen Zyklus (12, 13; 12+13=100).

Sample of Table S1: Timing and Duration of Specific Cycles
Coke Production, United States, 1914-1932

Dates of specific cycles	Timing at reference peak		Timing at reference trough		Duration of cyclical movements (mon.)						Per cent of duration of specific cycles	
	No. of months lead (-) or lag (+)	Date of reference peak	No. of months lead (-) or lag (+)	Date of reference trough	Specific cycles			Season over reference cycles				
					Expansion (6)	Contraction (7)	Full cycle (8)	Expansion (9)	Contraction (10)	Full cycle (11)	Expansion (12)	Contraction (13)
Trough - Peak - Trough (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Nov.14 - July 18 - May.19	-1	8/18	+1	12/14	44	10	54	0	+2	+2	81	19
May.19 - Aug.20 - July.21	+7	1/20	-2	9/21	15	11	26	+6	-9	-3	58	42
July.21 - May.23 - July.24	0	5/23	0	7/24	22	14	36	+2	0	+2	61	39
July.24 - Feb. 26 - Nov. 27	-8	10/26	-1	12/27	19	21	40	-8	+7	-1	48	52
Nov.27 - July 29 - Aug. 32	+1	6/29	-7	3/32	20	37	57	+2	-8	-6	35	65
Average	-0.2		-1.7		24.0	18.6	42.6	+0.4	-1.6	-1.2	57	43
Average deviation	3.4		1.9		8.0	8.3	10.3	3.5	5.5	2.6	12	12

Sample of Table S2: Amplitude of Specific Cycles
Coke Production, United States, 1914-1932

Dates of specific cycles	3-month average in specific-cycle relative centered on			Amplitude of			Per month amplitude of		
	Initial trough (2)	Peak (3)	Terminal trough (4)	Rise (5)	Fall (6)	Rise & fall (7)	Rise (8)	Fall (9)	Rise & fall (10)
Trough - Peak - Trough (1)									
Nov. 14 - July 18 - May 19	55.5	119.6	74.5	64.1	45.1	109.2	1.5	4.5	2.0
May 19 - Aug. 20 - July 21	88.7	125.9	45.0	37.2	80.9	118.1	2.5	7.4	4.5
July 21 - May 23 - July 24	44.3	144.0	82.6	99.7	61.4	161.1	4.5	4.4	4.5
July 24 - Feb. 26 - Nov. 27	69.4	118.2	92.1	48.8	26.1	74.9	2.6	1.2	1.9
Nov. 27 - July 29 - Aug. 32	105.5	141.6	41.7	36.1	99.9	136.0	1.8	2.7	2.4
Average	72.7	129.9	67.2	57.2	62.7	119.9	2.6	4.0	3.1
Average deviation	19.5	10.4	19.1	19.8	22.2	23.0	0.8	1.7	1.2

Sample of Table S3: Secular Movements
Coke Production, United States, 1914-1932

Dates of specific cycles	Average monthly standing (thous. short tons)			Per cent change from preceding phase		Per cent change from preceding cycle on base of			
	Expansion (2)	Contraction (3)	Full cycle (4)	Contraction to expansion (5)	Expansion to contraction (6)	Preceding cycle		Average of given and preceding cycle	
						Total (7)	Per month (8)	Total (9)	Per month (10)
Trough - Peak - Trough (1)									
Nov. 14 - July 18 - May 19	4193	4479	4246		+7				
May 19 - Aug. 20 - July 21	3906	3099	3565	-13	-21	-16	-0.40	-17	-0.42
July 21 - May 23 - July 24	3171	4326	3620	+2	+36	+2	+0.06	+2	+0.06
July 24 - Feb. 26 - Nov. 27	4107	4488	4307	-5	+9	+19	+0.50	+17	+0.45
Nov. 27 - July 29 - Aug. 32	4577	3319	3760	+2	-27	-13	-0.27	-14	-0.29
Average								-3.0	-0.05
Average deviation								12.5	0.30
Weighted average									-0.08

Sample of Table S4: Specific-cycle Patterns
Coke Production, United States, 1914-1932

Dates of specific cycles	Average in specific-cycle relative at stage										
	3 mos. centered on initial trough	Expansion				3 mos. centered on peak	Contraction				3 mos. centered on terminal trough
		First third	Middle third	Last third	First third		Middle third	Last third			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)		
Trough - Peak - Trough											
Nov. 14 - July 18 - May 19	55.5	81.3	108.0	107.1	119.6	118.7	112.6	88.2	74.5		
May 19 - Aug. 20 - July 21	88.7	101.2	110.9	117.8	125.9	124.4	86.8	50.4	45.0		
July 21 - May 23 - July 24	44.3	58.8	78.9	124.3	144.0	137.0	118.1	105.5	82.6		
July 24 - Feb. 26 - Nov. 27	69.4	84.8	95.1	106.2	118.2	108.3	105.2	99.0	92.1		
Nov. 27 - July 29 - Aug. 32	105.5	113.8	118.4	113.3	141.6	129.4	86.2	55.9	41.7		
Average	72.7	88.0	102.3	117.7	129.9	122.2	101.8	79.8	67.2		
Average deviation	19.5	15.1	12.2	8.9	10.4	6.9	12.2	2.1	19.1		

Tab. 3: Tabellen zur Charakterisierung von Zyklen nach A. F. Burns und W. C. Mitchell. Quelle: Burns/Mitchell (1946, S. 26-32).

Dates of specific cycles	Average change per month in specific-cycle relatives between stages									
	I-II	II-III	III-IV	IV-V	V-VI	VI-VII	VII-VIII	VIII-IX		
	Expansion					Contraction				
	Trough to first third (2)	First to middle third (3)	Middle to last third (4)	Last third to peak (5)	Peak to first third (6)	First to middle third (7)	Middle to last third (8)	Last third to trough (9)		
Trough - Peak - Trough (1)										
Nov. 14 - July 18 - May 19	+3.4	+1.8	-0.1	+1.7	-0.4	-2.0	-8.1	-6.8		
May 19 - Aug. 20 - July 21	+4.2	+2.2	+1.5	+2.7	-0.8	-10.7	-10.4	-2.7		
July 21 - May 23 - July 24	+3.6	+2.9	+6.5	+4.9	-2.8	-4.2	-2.8	-9.2		
July 24 - Feb. 26 - Nov. 27	+4.4	+1.7	+1.8	+3.4	-2.5	-0.5	-1.0	-1.7		
Nov. 27 - July 29 - Aug. 32	+2.4	+0.7	+2.3	+2.4	-3.0	-3.0	-2.5	-2.2		
Average	+3.6	+1.9	+2.4	+3.0	-1.9	-4.1	-5.0	-4.5		
Average deviation	0.6	0.6	1.6	0.9	1.0	2.7	3.4	2.8		
Average interval (mo.)	4.3	7.7	7.7	4.3	3.4	3.9	3.9	3.4		

Sample of Table R1: Reference-cycle Patterns
Coke Production, United States, 1914-1933

Dates of reference cycles	Average in reference-cycle relatives at stage										
	I	II	III	IV	V	VI	VII	VIII	IX		
	Expansion					Contraction					
	3 mos. covered on initial trough	First third	Middle third	Last third	3 mos. covered on peak	First third	Middle third	Last third	3 mos. covered on ter- minal trough		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)		
Trough - Peak - Trough											
Dec. 14 - Aug. 18 - Apr. 19	52.8	83.9	106.7	106.6	118.7	116.7	111.0	91.7	78.1		
Apr. 19 - Jan. 20 - Sep. 21	96.4	96.9	116.1	101.3	119.2	123.2	117.1	52.3	53.0		
Sep. 21 - May 23 - July 24	48.4	63.3	80.2	124.2	139.5	132.7	114.4	102.2	80.0		
July 24 - Oct. 26 - Dec. 27	69.5	90.0	98.4	111.4	107.4	105.0	102.4	94.5	94.0		
Dec. 27 - June 29 - Mar. 33	114.4	123.1	124.8	138.9	150.2	128.2	83.3	52.3	49.1		
Average	76.7	91.4	105.2	116.5	127.0	121.2	105.6	78.4	70.8		
Average deviation	23.8	14.8	12.8	12.0	14.3	8.2	10.2	21.0	15.8		

Sample of Table R2: Rate of Change from Stage to Stage of Reference Cycles
Coke Production, United States, 1914-1933

Dates of reference cycles	Average change per month in reference-cycle relatives between stages								
	I-II	II-III	III-IV	IV-V	V-VI	VI-VII	VII-VIII	VIII-IX	
	Expansion					Contraction			
Trough - Peak - Trough (1)	Trough to first middle third (2)	First to middle third (3)	Middle to last third (4)	Last third to peak (5)	Peak to first third (6)	First to middle third (7)	Middle to last third (8)	Last third to trough (9)	
Dec. 14 - Aug. 18 - Apr. 19	+4.1	+1.6	0.0	+1.6	-1.3	-2.3	-7.7	-9.1	
Apr. 19 - Jan. 20 - Sep. 21	-0.8	+7.7	-5.9	+9.0	+1.1	-0.9	-10.0	+0.2	
Sep. 21 - May 23 - July 24	+4.3	+2.6	+6.8	+4.4	-2.7	-4.1	-2.7	-6.9	
July 24 - Oct. 26 - Dec. 27	+4.1	+1.0	+1.6	-0.8	-1.0	-0.6	-1.8	-0.2	
Dec. 27 - June 29 - Mar. 33	+2.5	+0.3	+2.6	+3.2	-2.8	-3.1	-2.1	-0.4	
Average	+2.8	+2.6	+1.0	+3.5	-1.3	-2.2	-4.9	-3.7	
Average deviation	1.6	2.0	3.2	2.6	1.1	1.2	3.2	4.3	
Average interval (mo.)	4.3	7.5	7.5	4.3	3.6	6.5	6.5	3.6	

Sample of Table R3: Conformity to Business Cycles
Coke Production, United States, 1914-1933

Dates of reference cycles	Change in reference-cycle relatives during							Av. change per month during reference contraction minus that during	
	Reference expansion				Reference contraction			Preceding reference expansion*	Succeeding reference expansion*
	Total change in months (2)	Inter-val change in month (3)	Average change per month (4)	Total change (5)	Inter-val in months (6)	Average change per month (7)			
Trough-Peak-Trough (1)									
Dec. 14 - Aug. 18 - Apr. 19	+65.9	44.0	+1.50	-40.6	8.0	-9.08	-6.58	-	-
Apr. 19 - Jan. 20 - Sep. 21	+20.8	9.0	+2.31	-46.2	20.0	-2.31	-5.62	-	-
Sep. 21 - May 23 - July 24	+91.1	20.0	+4.56	-59.5	14.0	-4.55	-8.81	-	-
July 24 - Oct. 26 - Dec. 27	+37.9	27.0	+1.40	-13.4	14.0	-0.76	-2.36	-	-
Dec. 27 - June 29 - Mar. 33	+35.8	18.0	+1.99	-101.1	45.0	-2.25	-4.24	-	-
Average	+50.3	23.3	+1.52	-56.2	17.5	-3.17	-5.52	-	-
Average deviation	...	0.88	1.25	1.78	-	-
Index of conformity to reference									
Expansions	+100								
Contractions	+100								
Cycles, trough to trough	+100								
Cycles, peak to peak	+100								
Cycles, both ways	+100								

*Only the sign of the difference is entered.

Tab. 3 (Fortsetzung)

Der zweite Tabellentyp (S2) listet für jedes Zeitintervall des spezifischen Zyklus einer Reihe (1) den (Prozent-) ¹²⁹ Wert des Start-Tiefpunktes (2), des Hochpunktes (3) und des End-Tiefpunktes ($4=2+3$), anschließend die Differenz zwischen Hochpunkt und Tiefpunkten ($5=((3/2)-1)100$, $6=3-4$); Prozentpunkte Anstieg und Niedergang (die "Amplitude"), deren Summe ($7=5+6$) und schließlich diese Werte dividiert durch die Anzahl der Monate des zugrundeliegenden Zeitraums (die "Amplituden" pro Monat).

Der dritte Tabellentyp (S3) "throws into relief the secular component of the specific cycles": ¹³⁰ Für jeden Zeitraum eines spezifischen Zyklus (1) wird der Monatsdurchschnitt der absoluten Summe (z. B. einer Produktion) während der Expansionsphase (2), der Kontraktionsphase (3) und des gesamten Zeitraums des spezifischen Zyklus (4) tabelliert, weiterhin die prozentuale Veränderung der Produktion der Kontraktionsphase gegenüber der Expansionsphase (5) und der Expansionsphase gegenüber der Kontraktionsphase (6), schließlich die prozentualen Veränderungen gegenüber dem vorhergehenden Zyklus total (7) und pro Monat (8) auf der Basis des vorhergehenden Zyklus sowie auf der Basis des Durchschnitts beider Zyklen total (9) und pro Monat (10).

Der vierte Tabellentyp (S4) sollte die zyklischen Muster charakterisieren. In ihm finden sich für jeden Zeitraum eines spezifischen Zyklus (1) Durchschnitte der Relativwerte der spezifischen Zyklen für neun verschiedene Phasen des Zyklus. Tabellentyp S5 schließlich zeigt die Wachstumsrate eines spezifischen Zyklus von einem Stadium zum nächsten.

Tabellentyp R1 präsentiert das Muster der Referenzzyklen. Dieser Typus entspricht im Aufbau S4, wobei hier jedoch die Hoch- und Tiefpunkte der Referenzdaten anstelle der spezifischen Zyklen zugrundegelegt wurden und die Einheiten sich auf die Referenzzyklen beziehen. Tabellentyp R2 zeigt wie S5 die Wachstumsrate eines Zyklus von einem Stadium zum nächsten, nun jedoch für einen Referenzzyklus (dieser Typ geht aus R1 hervor wie S5 aus S4).

Die Vergleiche in S1 zwischen spezifischen und Referenzzyklen zeigten an, wie sich die zyklische Bewegung in einer Serie im Vergleich zur allgemeinen Konjunkturlage verhielt. Ein Vergleich zwischen dem durchschnittlichen spezifischen Zyklusmuster und dem Referenzzyklusmuster (S4 und R1) lieferte schließlich Tabellentyp R3. Für jede Referenzzyklusperiode (1) zeigt dieser Typ die Differenz, die durchschnittliche Länge in Monaten und die monatliche durchschnittliche Differenz von Tief- zu Hochpunkt während einer Expansionsphase (2, 3, 4) sowie während der anschließenden Kontraktionsphase (5, 6, 7). Zuletzt wurde noch die durchschnittliche monatliche Differenz während einer Kontraktion minus dessen Dauer im Vergleich zur vorhergehenden (8) bzw. nachfolgenden (9) Expansionsphase angegeben. Zur Illustration einzelner Werte konstruierten Burns und Mitchell zusätzlich graphische Darstellungen von zyklischen Mustern, die nach dem in Abb. 10 wiedergegebenen Schema aufgebaut wurden.

¹²⁹ Prozentual zum Mittelwert während des Zyklus, s. o. Die Relativ-Werte wurden zuvor mit einem dreigliedrigen gleitenden Mittelwert geglättet.

¹³⁰ Burns/Mitchell (1946, S. 28).

Sample Chart of Cyclical Patterns
(Drawn to twice the standard scales)

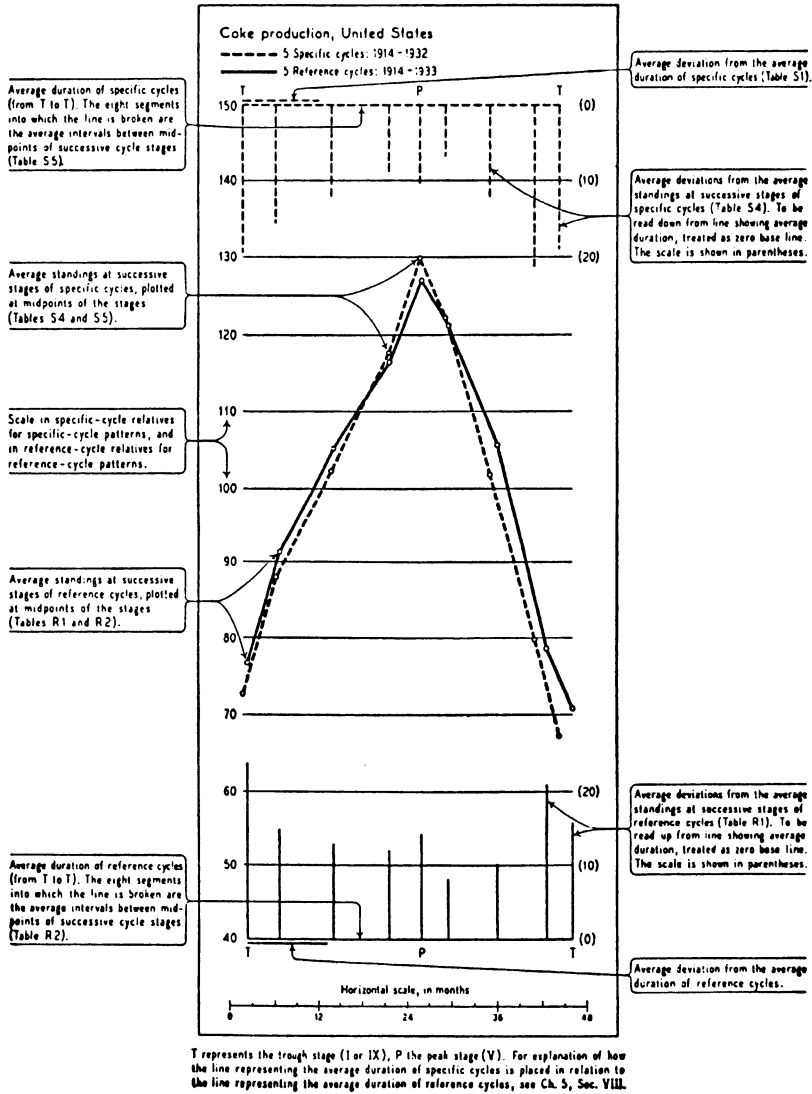


Abb. 10: Beispielgraphik für die Darstellung von Zyklennustern nach Burns und Mitchell.
Quelle: Burns/Mitchell (1946, S. 35).

Die monumentale Arbeit lieferte zunächst einmal ein immenses Zahlenmaterial. Zwar fanden sich vereinzelt auch Versuche, die empirischen Resultate über die Dauer der Zyklen oder eine säkulare Veränderung ihres Musters formal zu testen, der vorrangige Aspekt war jedoch die Zusammenstellung und Bewertung des Materials anhand der oben beschriebenen deskriptiven Kennzeichen. Zumindest die von Schumpeter aufgestellte Hypothese, daß jeweils drei Kitchin-Zyklen (40 Monate) in einem Juglar-Zyklus (9-10 Jahre) und sechs Juglar-Zyklen in einem Kondratieff-Zyklus enthalten waren, konnte so widerlegt werden.

Burns und Mitchell sahen auch diese Arbeit nur als Vorstufe für spätere Analysen. Die Autoren hofften, daß dann weitere Reihen – so etwa soziale Entwicklungen – anhand des entwickelten Schemas analysiert werden könnten.¹³¹ Diese Hoffnung sollte sich jedoch nicht erfüllen. Nach Mitchells Tod 1948 wurde vom NBER zwar der eingeschlagene Kurs beibehalten,¹³² doch in methodologischer Hinsicht dominierte zu dieser Zeit bereits eine Richtung, die sich selbst “ökonometrisch” nannte und der hier geschilderten Entwicklung ein theorieleses Vorgehen vorwarf. Diese Kritik erreichte nach Erscheinen des Werkes von Burns/Mitchell einen Höhepunkt in einer Diskussion zwischen T. C. Koopmans und R. Vining. Bevor wir uns dieser Kritik zuwenden, müssen wir jedoch zunächst die Gegenposition darstellen, die durch eine Verbindung von ökonomischer und statistischer Theorie die Analyse realwirtschaftlicher Phänomene explizit in einen stochastischen Rahmen stellen wollte.

3. Inferenz und Modellbildung

a. Allgemeines

Wir wollen das Thema dieses Kapitels anhand eines Beispiels von D. F. Hendry illustrieren.¹³³ Der ökonometrische Erkenntnisprozeß vollzieht sich nach Ansicht Hendrys auf vier Ebenen, die er mit dem in der Stochastik klassischen Beispiel des Würfels erläutert. Auf der ersten Ebene befindet sich die Wahrscheinlichkeitstheorie. Innerhalb dieser Ebene ist die Wahrscheinlichkeitsstruktur aller Ereignisse bekannt. Bezogen auf einen Wurf mit einem idealen Würfel ist sowohl das Modell (sechs mögliche, gleichwahrscheinliche, unabhängige Ausgänge) als auch die konkrete Wahrscheinlichkeit für das Ereignis “Sechs” eines Wurfes bekannt. Solch ein Wissensniveau sei in der Ökonometrie in der Regel unrealistisch (die Struktur ökonomischer Mechanismen sei größtenteils unbekannt), aber dennoch sei eine Beschäftigung mit Problemen auf dieser Ebene für ein weitergehendes Verständnis notwendig.

¹³¹ Aus diesem Grund hatte das NBER bereits ein Dutzend Monographien initiiert. Vgl. Burns/Mitchell (1946, S. 22).

¹³² Diese Vorgehensweise hat zumindest auf die wirtschaftspolitische Diskussion einen nicht unbedeutenden Einfluß ausgeübt. Zur Bedeutung der Datierungen in der weiteren Entwicklung des NBER siehe Moore/Zarnowitz (1986).

¹³³ Vgl. zum folgenden Hendry (1995, S. 16ff).

Auf der zweiten Ebene stehe ein spezifischer Würfel im Mittelpunkt des Interesses. Es sei nun lediglich bekannt, daß der Würfel sechs Seiten habe und die Ereignisse unabhängig seien. Das Ziel sei nun, die Wahrscheinlichkeit eines Sechserwurfes zu *schätzen*. A priori können wir hierüber keine genaue Aussage treffen, da wir nicht wissen, wie gleichmäßig dieser spezielle Würfel ist.¹³⁴ Wir können diese Wahrscheinlichkeit nur durch eine große Anzahl von Experimenten schätzen, wobei wir den relativen Anteil der Sechserwürfe als Schätzung für die feste "Populationswahrscheinlichkeit" verwenden. Diese Vorgehensweise entspräche der Schätzproblematik in der Ökonometrie, setze allerdings auf Ebene A getroffene Annahmen über das "Modell" voraus, die diese Schätzmethode rechtfertigen.

Doch auch diese Ebene sei von der Realität noch weit entfernt. Die Annahme, daß der Würfel während des Würfels seine Form nicht verändere (sechs Seiten mit unterschiedlichen Zahlen etc.) sei für die Ökonomie kaum zutreffend.¹³⁵ Hier liege eher eine Situation vor, die er mit Ebene C bezeichnet.

Auf dieser Ebene C sei lediglich bekannt, daß ein Würfel mit Zahlen existiere sowie die Ergebnisse einer gewissen Anzahl von durchgeführten Würfeln, anhand derer die Wahrscheinlichkeit, einen Sechserwurf zu beobachten, ermittelt werden solle. Wir wissen nicht, wieviele Seiten der Würfel hat, auf wievielen Seiten sich eine Sechs befindet, ob er gleichmäßig ist und die einzelnen Würfe unabhängig voneinander sind, ob sie überhaupt jedes Mal auf dieselbe Art und Weise durchgeführt wurden etc. Die Wahrscheinlichkeit eines "Sechserwurfes" könne nun nur ermittelt werden, wenn man anhand der beobachteten Würfe ein Modell des Generierungsmechanismus konstruiere und anhand dieses Modells die Wahrscheinlichkeit bestimmt werde. Diese Situation entspräche der typischen Problematik der empirischen Ökonometrie. Hier sei in der Regel eine Vielzahl von stochastischen (dynamischen) Modellen realistisch, die gegeneinander abgewogen werden müßten. Die vierte Ebene schließlich bestehe aus dem Problem der Prognose zukünftiger oder alternativer Ereignisse anhand des konstruierten Modells und der daraus abgeleiteten Parameter.

Hendry betont zu Recht, daß sich die "klassische" Ökonometrie der 50er und 60er Jahre zum überwiegenden Teil auf den Ebenen A und B bewegt habe¹³⁶ und vor allem Ebene C mehr Beachtung geschenkt werden sollte.¹³⁷ Da sich aufgrund der komplexen Natur des Gegenstandes und unseres vergleichsweise geringen Wissensstandes ein logisches Vorgehen von A über B und C zu D nicht realisieren lasse, sei hierbei ein *iteratives* Vorgehen notwendig.¹³⁸

¹³⁴ Hierbei handelt es sich offensichtlich um die Situation, die Bernoulli vor Augen hatte. Siehe oben, Abschnitt A 3 b.

¹³⁵ Hendry spricht hier davon, daß die *Annahme* eines konstanten Würfels mit einem Wissensniveau korrespondiere, das in der Ökonomie nicht erreicht werden könne. Die Frage, die sich hierbei aufdrängt, ist, weshalb man dies überhaupt annehmen sollte.

¹³⁶ "[...] they mainly analyse mathematical-statistical problems, not empirical discovery problems." Hendry (1995, S. 18).

¹³⁷ Die von Hendry behandelten Ebenen A, B und C umfassen – in den Größenordnungen von etwa 40 %, 30 % und 30 % – immerhin 869 Seiten, Ebene D ist ein eigenes Buch gewidmet. In den klassischen Ökonometrie-Lehrbüchern der 60er und 70er Jahre habe Ebene B 60-70%, Ebene A etwa 10-20 % und die Ebenen C und D nur einen geringen Teil ausgemacht.

Analog zu den vorhergehenden Kapiteln wollen wir auch hier nicht versuchen, einen systematischen Zugang zum Verständnis dieser Ebenen zu öffnen, sondern wiederum die historische Entwicklung ihrer einzelnen Komponenten und des jeweiligen Zusammenspiels in den Mittelpunkt stellen.

Wir werden zuerst die Anfänge "ökonometrischer" Traditionen behandeln, die formell mit der Gründung der *Econometric Society* 1929 und dem ersten Erscheinen des von dieser Gesellschaft herausgegebenen Publikationsorgans *Econometrica* 1933 angesetzt werden kann. Vor allem seit der Untersuchung von Morgan (1990), deren Einschätzung mittlerweile allgemein gefolgt wird, wird die inferenzstatistische Fundierung der Ökonometrie in dem von ihr als *probability revolution* bezeichneten Beitrag von Haavelmo (1944) gesehen. In der Tat hat diese Arbeit der nachfolgenden ökonometrischen Forschung eine Richtung gewiesen, der sie für Jahrzehnte fast ohne Ausnahme folgte.

Wie jede Revolution hat auch die Haavelmosche Voraussetzungen. Neben den noch zu diskutierenden Adaptionen der theoretischen Statistik ist der ökonometrische Weg dorthin vor allem durch drei Namen gekennzeichnet: durch Ragnar Frisch, Tjalling Koopmans und Jan Tinbergen. Alle drei, die man als Gründungsväter der Ökonometrie bezeichnen kann, gingen über ihre – im vorigen Abschnitt besprochenen – Vorläufer vor allem in einem Punkt hinaus: sie verwendeten explizit probabilistische Ansätze. Frisch verband Yules Ansatz stochastischer Zyklen mit einem ökonomischen theoretischen Ansatz und stellte, wie wir gesehen hatten, ebenso wie dieser und Slutsky die "Schocks" selbst in den Mittelpunkt, allerdings ohne die Parameter zu schätzen. Weiterhin sah er den Grund für Diskrepanzen zwischen theoretischen Größen und realen Daten in "Fehlern in den Variablen". Koopmans fragte zusätzlich noch nach dem "sampling error", der ebenfalls berücksichtigt werden müsse. Diesen vorrangig theoretischen Arbeiten folgte kurz darauf die umfassende Analyse von Tinbergen, in der er im Auftrag der Vereinten Nationen Konjunkturzyklen untersuchte, im Gegensatz etwa zur Tradition des NBER oder der im selben Jahr erschienenen Arbeit von Schumpeter jedoch in einem probabilistischen Rahmen. Wie die bereits erwähnte monumentale Studie von Burns und Mitchell in der berühmten "Measurement-without-theory"-Debatte Gegenstand kritischer Einwände seitens der Ökonometriker wurde, so fand Tinbergens Ansatz in einem nicht minder berühmten Disput in J. M. Keynes einen vehementen Gegner, der sich gegen die probabilistischen Grundannahmen wandte. Dieser Streit konnte dem einsetzenden Siegeszug der Ökonometrie keinen lebensbedrohenden Schlag mehr versetzen. Es ist sogar, "Ironie der Wissenschaftsgeschichte"¹³⁹, gerade die keynesianische Theorie gewesen, die in der Folgezeit ökonometrischen Modellen zum Aufschwung verhalf.

¹³⁸ Vgl. Hendry (1995, S. 19).

¹³⁹ Kirchgässner (1983, S. 513).

b. Anfänge ökonometrischer Traditionen

Wenn wir uns nun den Anfängen ökonometrischer Traditionen zuwenden, so ist zunächst zu fragen, warum diese erst in den dreißiger Jahren des 20. Jahrhunderts einsetzten. Warum gab es nicht schon früher systematische probabilistische Ansätze? Dies ist insofern erstaunlich, als eine Reihe von ökonomischen Theoretikern wie Cournot oder Walras sich auch mit wahrscheinlichkeitstheoretischen Problemen auseinandergesetzt haben.¹⁴⁰ Keiner dieser Theoretiker sah jedoch eine Verbindung zwischen der Wirtschafts- und Wahrscheinlichkeitstheorie. Wie wir in Abschnitt A 3 c gesehen hatten, war F. Y. Edgeworth maßgeblich an der Entwicklung der theoretischen Statistik gegen Ende des 19. Jahrhunderts beteiligt und verwendete dabei auch eine Reihe von sozioökonomischen Daten. Doch selbst Edgeworth unterschied strikt zwischen statistischen und (wirtschafts-) theoretischen Untersuchungen.¹⁴¹

Die von Edgeworth, Pearson und Yule entwickelten Methoden wie die Korrelations- und Regressionsrechnung fanden zwar seit dem ersten Jahrzehnt des zwanzigsten Jahrhunderts auch in statistischen Untersuchungen in der Ökonomie, z. B. in Nachfrageanalysen, Verwendung, doch geschah dies eher in deskriptivem Sinne. Und auch die Analyse von Zeitreihen wurde, wie wir im letzten Abschnitt gesehen hatten, nicht im Rahmen einer stochastischen Konzeption betrieben. Zum einen hatte hier eine Koryphäe der theoretischen Statistik (Yule) vor "nonsense correlations" gewarnt, zum anderen wurde gerade bei Zeitreihen die Strukturkonstanz, Voraussetzung für die Verwendung von Verteilungen von Koeffizienten, als problematisch angesehen.¹⁴²

Nach M. Morgan war man zu dieser Zeit allgemein der Auffassung, daß ökonomische Daten nicht die Anforderungen der Wahrscheinlichkeitstheorie erfüllten.¹⁴³ Abweichungen zwischen einem Modell und gegebenen Daten wurden zum einen in der Tradition des *ceteris-paribus*-Denkens mit unberücksichtigten Faktoren erklärt, zum anderen mit dem statistischen Argument der Meßfehler.

Es finden sich aber auch vereinzelt Ansätze, die die Anwendung wahrscheinlichkeitstheoretischer Konzepte bei der quantitativen Analyse ökonomischer Daten für sinnvoll hielten. 1924 widmete sich W. C. Mitchell in einer *Presidential Address* vor der *American Economic Association* dem Problem der quantitativen Analyse und ihrer Beziehung zur ökonomischen Theorie.¹⁴⁴ Mitchell kritisierte, daß die ökonomische

¹⁴⁰ Vgl. Menard (1987). Menard wirft allerdings nur die Frage auf, weshalb die Wahrscheinlichkeitstheorie keinen Einfluß auf die ökonomische Theorie hatte und geht nicht auf die Anwendung in bezug auf ökonomische Daten ein.

¹⁴¹ Vgl. Porter (1986, S. 268f), Menard (1987, S. 141). Überraschenderweise wird Edgeworth in diesem Zusammenhang weder von Morgan (1987) noch von Morgan (1990), Qin (1993) oder Hendry/Morgan (1995b) erwähnt.

¹⁴² Epstein (1987, S. 52 Anm. 8) bemerkt hierzu: "The neglect of standard errors was typical in this era, possibly because sample sizes were small and applied workers were not yet familiar with the *t* distribution that had been derived some years earlier." Diese Interpretation scheint uns jedoch zu kurz zu greifen.

¹⁴³ Vgl. Morgan (1990, S. 230ff). Siehe auch Morgan (1987).

¹⁴⁴ Vgl. die schriftliche Fassung, Mitchell (1925).

mische Theorie noch immer zu sehr dem mechanischen Denken der *traditionellen* Physik mit der Annahme von unveränderlichen Gesetzen verpflichtet sei, andererseits in der realen Welt jedoch kein Faktor konstant und immerwährend wirke. Hier solle man sich an der *modernen* Physik orientieren: "The difference between the mechanical and the statistical conceptions of nature has been clearly worked out in physics. The mechanical view involves the notions of sameness, of certainty, of invariant laws; the statistical view involves the notions of variety, of probability, of approximations."¹⁴⁵ In der Ökonomie habe man es im Vergleich zur Physik bei einer quantitativen Analyse mit größeren Problemen zu tun: Die Anzahl der Daten sei geringer, die Strukturen veränderten sich, und man könne keine Experimente durchführen. Dennoch sollte auch die ökonomische Theorie den Fakten mittels dieser statistischen Denkweise angepaßt werden.

Zur gleichen Zeit ging W. M. Persons in einem umfangreichen Beitrag auf die Beziehung zwischen Statistik und ökonomischer Theorie ein.¹⁴⁶ Er war der Ansicht, daß es durchaus möglich sei, einen großen Teil der im Zusammenhang mit der Wahrscheinlichkeitstheorie entwickelten statistischen Techniken auf die Ökonomie zu übertragen.¹⁴⁷ Persons schilderte ausführlich die auf Bayes und Laplace aufbauende Entwicklung der *inversen* Wahrscheinlichkeitstheorie (s. o., Abschnitt A 2 b), deren fruchtbare Anwendungsmöglichkeiten in der Ökonomie bereits von F. Y. Edgeworth aufgezeigt worden seien. Auch andere führende theoretische Ökonomen wie W. S. Jevons, W. Bagehot, J. N. Keynes, A. Marshall und sogar J. S. Mill hätten die Bedeutung statistisch-induktiver Untersuchungen als Komplement zur deduktiven Vorgehensweise der theoretischen Ökonomie anerkannt. Persons blickte optimistisch in die Zukunft:

"The statistical work of Moore, Mitchell, [Irving] Fisher, Taussig, and others in part verifies theory, in part suggests revisions of statement, and in part suggests new laws. Statistical investigations of the future, as in the past, may be expected to start either with current economic theory or some empirical observation as a 'first approximation' and proceed to verification or amendment of economic law. Statistical description and statistical inference, then may be expected to make economic theories more realistic, and in general give us more certain and useful answers to economic questions [...]"¹⁴⁸

Andererseits wollte Persons den Bereich der Analyse von Zeitreihen ausgeschlossen wissen. Ein Jahr zuvor hatte er, immerhin als Präsident der *American Statistical Association*, festgestellt: "Granting as he [der Statistiker, T. R.] must, that consecutive items of a statistical time series are, in fact, related, he admits that the mathematical theory of probability is inapplicable."¹⁴⁹

¹⁴⁵ Ebda., S. 11.

¹⁴⁶ Diese Ausführungen werden von Morgan (1990) überraschenderweise nicht behandelt.

¹⁴⁷ Vgl. Persons (1925, S. 182).

¹⁴⁸ Ebda., S. 197. Die Arbeiten von Moore und Mitchell hatten wir bereits im letzten Abschnitt angesprochen. I. Fisher und F. W. Taussig hatten sich ebenfalls in mehreren Zusammenhängen quantitativen Analysen gewidmet.

¹⁴⁹ Persons (1924b, S. 11), zitiert nach Darnell (1994b, S. xv).

Den vehementesten Angriff auf die Verwendung stochastischer Ansätze in der Ökonomie lieferte 1932 L. Robbins. Robbins bestritt, daß Untersuchungen von "random samples" überhaupt zu Verallgemeinerungen führen, geschweige denn ökonomische Daten überhaupt als solche "random samples" aufgefaßt werden könnten.¹⁵⁰ Wenn man etwa die Nachfrageelastizität für Hering für den Zeitraum 1907-1908 numerisch mit 1.3 bestimmt habe, könne damit noch lange kein Gesetz postuliert werden, denn dieser Wert hänge von zu vielen Faktoren ab: von Geschmacksänderungen, der Verfügbarkeit anderer Produkte, der Einkommensverteilung etc., vor allem jedoch von Faktoren, die sich im Laufe der Zeit veränderten, wie etwa technologische Entwicklungen oder Transportkosten. Er fragte daher: "Is it possible reasonably to suppose that coefficients derived from the observation of a particular herring market at a particular time and place have any *permanent* significance – save as Economic History?"¹⁵¹

Wenn man die Beobachtung der Angebots- und Preisentwicklung über einen längeren Zeitraum ausdehne, so könne man zwar Zahlen erhalten, die die durchschnittliche Elastizität während einer längeren Periode wiedergeben. Diese Zahlen hätten durchaus einen Erkenntniswert, allerdings im deskriptiven Sinne:

"They are a convenient way of describing certain forces operative during that period of history. As we shall see later on, they may provide some guidance concerning what may happen in the immediate future. But they have no claim to be regarded as 'laws'. However accurately they describe the past, there is no presumption that they will describe the future. [...] Important as such investigations may be, at the moment at which they are made and perhaps for a short time after, there is no justification for claiming for their results the status of the so-called 'statistical' laws of the natural sciences."¹⁵²

Eine solche Berechtigung sahen andere jedoch durchaus. Diese geradezu diametrale Gegenposition ist mit dem Namen der *Econometric Society* verbunden, die den Startpunkt der "offiziellen" Entwicklung der Ökonometrie markiert.

Bereits 1912 hatte I. Fisher versucht, eine Gesellschaft zur Förderung einer quantitativen und mathematisch orientierten Ökonomie zu gründen.¹⁵³ Zwar hatten unter anderem W. C. Mitchell und H. L. Moore Interesse bekundet, doch war dem Vorstoß kein Erfolg beschieden. Erst 1928 wurde die Idee wieder aufgegriffen, als R. Frisch ein Jahr in den USA verbrachte und in Princeton mit dem jungen Mathematiker C. F. Roos zusammentraf. Roos war zu diesem Zeitpunkt Sekretär der Sektion "Ökonomie, Soziologie und Statistik" der *American Association for the Advancement of Science*. Beide waren sich einig, daß Mathematik und Statistik eine größere Rolle in der Ökonomie spielen sollten. Sie nahmen Kontakt zu I. Fisher auf, und anläßlich einer gemeinsamen Tagung der *American Economic Association*, der *American Statistical Association* und der Sektion "Ökonomie, Soziologie und Statistik" der *American Association for the Advancement of Science* im Dezember 1930 in Cleveland wurde von Roos zu einer Tagung einer *Econometric*

¹⁵⁰ Vgl. Robbins (1932, S. 99).

¹⁵¹ Ebda., S. 100. Hervorhebung im Original.

¹⁵² Ebda., S. 101.

¹⁵³ Vgl. zum folgenden Christ (1983).

Society eingeladen,¹⁵⁴ zu der zwölf Amerikaner und vier Europäer erschienen. J. A. Schumpeter, zu dieser Zeit noch Professor in Bonn, wurde zum Tagungsvorsitzenden gewählt, I. Fisher zum ersten Präsidenten. Die erste Fassung einer Satzung wurde von Frisch vorgeschlagen und anschließend von F. C. Mills, C. Roos und ihm selbst ausformuliert.¹⁵⁵ Einige Passagen daraus finden sich im Editorial der ersten Ausgabe.¹⁵⁶ Das Hauptziel der Gesellschaft lautete: "Its main object shall be to promote studies that aim at a unification of the theoretical-quantitative and the empirical-quantitative approach to economic problems and that are penetrated by constructive and rigorous thinking *similar to that which has become to dominate in the natural sciences*."¹⁵⁷ Insbesondere einer Mathematisierung wurden dabei die Türen geöffnet. Die Artikel der Zeitschrift sollten ein möglichst hohes mathematisches Niveau haben: "[...] it will be an editorial principle of *Econometrica* that no paper shall be rejected solely on the ground of being too mathematical. This applies no matter how highly involved the mathematical apparatus may be."¹⁵⁸ Der Eröffnungsartikel von J. Schumpeter war von großem Enthusiasmus geprägt. Er sprach hier von einem "Pantheon" der Ökonometrie, der über W. S. Jevons, J. H. von Thünen, A. A. Cournot bis zu G. King und W. Petty zurückreichte.¹⁵⁹

¹⁵⁴ Der Name wurde in Anlehnung an die englische "biometrische Schule" von K. Pearson gewählt.

¹⁵⁵ Nach zwei abgehaltenen Tagungen in Lausanne und Washington 1931 wuchs die Mitgliederzahl (jeder Tagungsteilnehmer wurde Mitglied) rasch auf 173 Personen an. Durch ein großzügiges Angebot von Alfred Cowles wurde bereits zu Beginn eine dauerhafte finanzielle Absicherung geschaffen. A. Cowles von der Investmentgesellschaft *Cowles and Company* veröffentlichte in den zwanziger Jahren in Colorado Springs regelmäßig Anlagetips. Nach dem Börsenkrach von 1929, der die Prognosefähigkeit der Investmentgesellschaften stark erschütterte, stellte er 1931 die Publikation seiner Prognosen mit der Begründung ein, daß er die Mechanismen der Konjunkturzyklen und Aktienkursbewegungen nicht genügend verstehe, um solide Prognosen treffen zu können und er sich daher tiefergehend dem Studium dieser Mechanismen zuwenden wolle, bevor er weitere Prognosen anstelle. Cowles besprach dieses Vorhaben mit einem Bekannten (einem mathematisch ausgebildeten Biochemiker), der ihm die multiple Korrelationsrechnung als Hilfsmittel vorschlug und ihn seinerseits an einen Bekannten, den Mathematiker H. T. Davis, verwies. Cowles fragte Davis, ob es möglich sei, einen multiplen Korrelationskoeffizienten von 24 Variablen zu berechnen. Davis antwortete, er könne sich zwar nicht vorstellen, weshalb jemand so etwas tun wolle, daß es aber (mit den neuen, von IBM entwickelten Hollerith-Maschinen) prinzipiell möglich sei und verwies Cowles wiederum an die neugegründete *Econometric Society*. Dort seien Wissenschaftler zu finden, die auch langfristig solche Berechnungen für ihn durchführen könnten. Cowles schrieb daraufhin Fisher, der dem Angebot einer bezahlten Auftragsarbeit (zunächst \$ 12.000 pro Jahr) nach Absprache mit C. Roos erfreut zustimmen wollte. Die europäischen Mitglieder der Gesellschaft fürchteten um ihre wissenschaftliche Freiheit und entschieden, Frisch als Vertreter zu entsenden, der mit Cowles Kontakt aufnehmen sollte. Dieses Treffen verlief positiv, und man beschloß, Cowles Angebot anzunehmen. Gleichzeitig einigte man sich, eine Zeitschrift mit dem Namen *Econometrica* (der Name in Anlehnung an *Biometrika*) zu gründen. Frisch wurde der geschäftsführende Herausgeber, Cowles Schatzmeister der Gesellschaft. Somit hatte von Anfang an der Prognoseaspekt eine fundamentale Bedeutung hinter der epistemologische Intentionen zurückstanden.

¹⁵⁶ Vgl. Frisch (1933b).

¹⁵⁷ Ebda., S. 1. Hervorhebung von uns.

¹⁵⁸ Ebda., S. 3.

¹⁵⁹ Vgl. Schumpeter (1933, S. 6-9).

Wie wurde das "Programm" nun realisiert? Nach Hendry und Morgan waren die dreißiger Jahre in bezug auf die sich entwickelnde Ökonometrie durch den Versuch gekennzeichnet, die von R. A. Fisher für die Naturwissenschaften, speziell für die Biologie, entwickelten multivariaten Analyseverfahren zu übernehmen und für die Anwendung auf ökonomische Daten fruchtbar zu machen.¹⁶⁰ Dabei waren vor allem zwei besondere Aspekte zu berücksichtigen. Ökonomische Daten waren zum einen passive Daten, konnten also in der Regel nicht in kontrollierten Experimenten erzeugt werden und lagen zum anderen in der Regel in Form von Zeitreihen vor. Außerdem existierten zumeist mehrere Beziehungen *gleichzeitig*.¹⁶¹ In den Naturwissenschaften war es möglich, durch das "Design" des Experimentes zu gewährleisten, daß die Annahmen, die eine optimale Schätzung ermöglichten, zutrafen. In der Ökonomie konnte man diese Annahmen allenfalls prüfen. Waren die Bedingungen nicht erfüllt, mußten spezifische Lösungen gefunden werden. Entsprechende Lösungsversuche wurden im folgenden in vier Schritten von R. Frisch, T. Koopmans, J. Tinbergen und T. Haavelmo formuliert. Wenden wir uns zunächst R. Frisch zu.

Frisch hatte sich, wie in Abschnitt B 2 c erwähnt, zunächst mit dem Problem der Auswirkungen von "Schocks" in einem dynamischen System beschäftigt. Dabei spielten stochastische Überlegungen keine Rolle. Die Werte der Parameter wurden nicht anhand von realen Daten geschätzt, geschweige denn Aussagen über ihre Präzision getroffen, sondern statt dessen in einem *trial-and-error*-Verfahren so bestimmt, daß das System eine Entwicklung nahm, die der Realität nahe kam.

Frisch widmete sich jedoch auch dem Problem der *Schätzung* von Parametern. In einer ein Jahr nach seinem berühmten *Propagation*-Aufsatz erschienenen Monographie *Statistical Confluence Analysis by Means of Complete Regression Systems* versuchte er, eine Methodologie zu entwickeln, die den spezifischen Problemen der Ökonomie angemessen sein sollte. Er erkannte, daß eine einfache Übertragung der für die Naturwissenschaften entwickelten statistischen Verfahren auf die Ökonomie nicht möglich war. Zum einen seien hier alle Variablen mit Meßfehlern behaftet, zum anderen existiere zwischen ökonomischen Variablen eine "Konfluenz".¹⁶² Vor allem die Berücksichtigung des letzteren stellte er in den Mittelpunkt seiner Forschungen. Nach Hendry und Morgan ging es ihm dabei um nicht weniger als eine zu R. A. Fishers Ansatz kongeniale Fundierung der statistischen Analyse in der Ökonomie.¹⁶³ Ein erheblicher – und aus Sicht der Ökonomie verständlicher – Nachteil des Fisherschen Ansatzes war, daß er keine statistischen Kriterien für die

¹⁶⁰ Vgl. Hendry/Morgan (1995b, S. 33).

¹⁶¹ Diese Gleichzeitigkeit ist zunächst unspezifisch im Sinne von wechselseitigen Beziehungen ohne eine genaue Spezifikation der zeitlichen Abfolge gemeint.

¹⁶² Dies würde man heute als Kollinearität bezeichnen. Die erste explizite Unterscheidung der Vorstellung, daß Abweichungen zwischen einem Modell und Daten "Meßfehler" in den Variablen sind ("errors-in-variables"), von der Vorstellung, daß sie auf unberücksichtigten Variablen beruhen ("errors-in-equations"), geht, zumindest im angelsächsischen Raum, auf Schultz (1928) zurück.

¹⁶³ Vgl. Hendry/Morgan (1995b, S. 40f). Sie betonen (ebda.), daß der Ansatz von Frisch einer wahrscheinlichkeitstheoretischen Fundierung ermangelte.

Auswahl der einzubeziehenden Variablen oder die Form des Modells vorsah. In den von Fisher zugrundegelegten Anwendungsbereichen der experimentellen Naturwissenschaften waren diese Fragen auch nicht von Interesse; dort waren die relevanten Größen durch das "Design of Experiments" vorgegeben.¹⁶⁴ Wenn man nun die wechselseitigen Beziehungen zwischen den Variablen nicht berücksichtigte und die *sampling theory* anwendete, waren die Ergebnisse nach Frischs Ansicht sinnlos. Die Stichprobentheorie sei zwar allgemein sinnvoll, aber für die Analyse ökonomischer (konfluenter) Daten noch nicht genügend ausgebaut. Diese Aufgabe lohne auch nicht den nötigen hohen Aufwand:

"Indeed, if the sampling aspect of the problem should be studied from a sufficiently general set of assumptions, I found that it would lead to such complicated mathematics that I doubted whether anything useful would come out of it. And, on the other hand, if the sampling aspect should be studied under simple assumptions, for instance, of not collinear and normally distributed basic variates, the essence of the confluence problem would not be laid bare."¹⁶⁵

Frisch lehnte hier also die Stichprobentheorie nicht aus grundsätzlichen Erwägungen heraus ab, sondern weil sie technisch noch nicht so weit entwickelt sei, um die spezifischen Probleme der Ökonomie berücksichtigen zu können. Er erkannte klar, daß die Verteilungstheorie nicht mehr anwendbar war, wenn man nicht mit Sicherheit von einem korrekten Modell ausgehen konnte und Variablen eines Modells per "judgement" hinein- oder herausnahm. Sein Lösungsvorschlag bestand in der Konstruktion von sog. "bunch maps", einer Reihe von graphischen Darstellungen, die die relative Bedeutung von Regressionskoeffizienten für alle möglichen Kombinationen der betrachteten Variablen anzeigen sollten.¹⁶⁶ Mit diesem Verfahren wollte Frisch der Ökonometrie ein komplettes Analyseinstrument in die Hand geben, daß die spezifischen Probleme ökonomischer Daten berücksichtigen sollte: das Problem der Multikollinearität, der Variablenauswahl und der Wahl der Modellform.¹⁶⁷

Der nächste Schritt zu einer systematischen ökonometrischen Methodologie wurde von T. C. Koopmans vollzogen.¹⁶⁸ Koopmans, der vor allem als Initiator der

¹⁶⁴ So der Titel des wohl berühmtesten Lehrbuches zu diesem Thema von Fisher. In der herkömmlichen Theorie waren die Regressoren feste Größen und keine Variablen.

¹⁶⁵ Frisch (1934, S. 7).

¹⁶⁶ Auf die – relativ komplizierten – technischen Einzelheiten wollen wir hier nicht eingehen. Siehe dazu Hendry/Morgan (1989). Für drei Variablen x_1, x_2, x_3 gibt es die folgenden Regressionsmöglichkeiten: $x_1 = b_{12.3}x_2 + b_{13.2}x_3, x_2 = b_{21.3}x_1 + b_{23.1}x_3, x_3 = b_{31.2}x_1 + b_{32.1}x_2$.

¹⁶⁷ Eine grundsätzliche Ablehnung schien er jedoch hinsichtlich der Anwendung der *sampling theory* in der Zeitreihenanalyse zu hegen. Dies legt jedenfalls eine Bemerkung H. O. A. Wolds nahe. Frisch war Gutachter seiner Dissertation (Wold (1938)), und Wold bemerkte hierzu: "He did not like my thesis, so he was very angry at me and tried to find reasons to show that I was wrong. He had a big point and was very fierce, but I insisted on my point and gradually had my way. [...] My understanding now is that he had a wrong feeling about what happened. It was a mistake for him to think this way, but he did." Wold in Hendry/Morgan (1994, S. 422).

¹⁶⁸ Tjalling Charles Koopmans [1910-1985] studierte bis 1933 Mathematik und Physik an der Universität von Utrecht und promovierte 1936 mit einer (wirtschaftswissenschaftlichen)

“Measurement-without-Theory”-Debatte in die Geschichte der Ökonometrie eingehen sollte (s. u.), erweiterte 1937 den Ansatz von Frisch. Seine Position unterscheidet sich jedoch auch in mehreren Aspekten von Frischs Ansatz, insbesondere hinsichtlich der zentralen Frage der Inferenzsicht.¹⁶⁹

Das Ziel seiner Untersuchung war wie bereits bei Frisch die Entwicklung einer allgemeinen Methodologie für die Analyse ökonomischer Daten. Im Gegensatz zu Frisch sah er jedoch mehr Variationsquellen. Er berief sich explizit auf R. A. Fishers Konzeption einer “hypothetisch infiniten Population”, eines “Universums”, von dem die beobachteten Daten als Ziehung einer Zufallsstichprobe angesehen werden könnten.¹⁷⁰ Für Frisch waren nur die *Messungen* der Variablen fehlerbehaftet, die Variablen selbst waren nicht (Zufalls-) Realisationen einer übergeordneten Verteilung, sondern einmalig. Koopmans ging dagegen von einem “repeated-sampling”-Ansatz aus. Dieser war natürlich rein hypothetisch:

“[...] a repeated sample consists of a set of values which the variables would have assumed if in these years the systematic components had been the same and the erratic components had been other independent random drawings from the distribution they are supposed to have.”¹⁷¹

Damit sind nicht nur die “wahren” Werte der Parameter hypothetische Größen, sondern auch die Verteilungen der Fehler. Nur so konnten ökonomische Daten, die einmalige Realisationen waren, mit Fishers Konzeption in Einklang gebracht werden.¹⁷²

Ein weiterer Unterschied bestand in der *Art* der Fehler. Koopmans ging davon aus, daß es in der Realität nicht gelingen werde, alle Regressoren zu ermitteln. Diese durch unberücksichtigte relevante Regressoren verursachten Fehler seien erheblich größer als die von Frisch alleine berücksichtigten Meßfehler.

Ein solches Programm erforderte ausgefeiltere Methoden als die bislang in der Ökonomie verwendeten. Koopmans wollte daher vor allem den Maximum-Likelihood-Ansatz von Fisher adaptieren, insbesondere die exakte Verteilung der Schätzer in einer Differenzgleichung ermitteln, um anhand der von Fisher entwickelten Schätztheorie sicherere Aussagen über makroökonomische Strukturen treffen zu können.¹⁷³ Um die mit dieser Methode verbundenen starken Annah-

Dissertation, *Linear Regression Analysis of Economic Time Series*, in Leiden bei J. Tinbergen. 1936 bis 1938 lehrte er an der Universität Rotterdam, anschließend arbeitete er bis 1940 in der Finanzabteilung der Vereinten Nationen in Genf. 1940 ging er in die Vereinigten Staaten und wurde dort 1944 Mitglied der *Cowles Commission* (s. u.), war von 1948 bis 1954 ihr Direktor, schließlich Professor in Yale. 1950 war er Präsident der *Econometric Society*, 1966 Vizepräsident der *American Economic Association*. 1975 erhielt er zusammen mit L. V. Kantorovich den Nobelpreis für Wirtschaftswissenschaften.

¹⁶⁹ Vgl. zum folgenden auch Morgan (1990, S. 238-240), Qin (1993, S. 16-18) sowie Hendry/Morgan (1995b, S. 41-43).

¹⁷⁰ Insbesondere auf den von uns in Abschnitt A 3 d erwähnten, grundlegenden Beitrag, Fisher (1922).

¹⁷¹ Koopmans (1937, S. 7).

¹⁷² Vgl. Hendry/Morgan (1995b, S. 42).

¹⁷³ Er berief sich hier auf Fisher (1925). “The point was that the likelihood framework made possible an entirely new battery of critical tests that could extract more information from the

men rechtfertigen zu können, führte Koopmans eine Reihe von Tests ein, so unter anderem die Überprüfung der Residuen, die unkorreliert und normalverteilt sein mußten.

Koopmans Ausführungen zeigten zunächst keine große Wirkung in der Ökonometrie.¹⁷⁴ Erst die Rezeption und konkrete Anwendung seiner und Frischs Ansätze durch J. Tinbergen verhalf einer stochastischen Denkweise zum Durchbruch. Koopmans selbst sollte in der "Measurement-without-Theory"-Debatte noch einmal auf die Frage der Modellbildung zurückkommen. Wenden wir uns jedoch zunächst Tinbergen zu.

c. Noch einmal Konjunkturzyklen: Die Tinbergen-Keynes-Kontroverse

Die Arbeiten von Frisch und Koopmans hatten versucht, der Ökonometrie eine inferenzstatistische Grundlage zu geben. Sie waren jedoch hauptsächlich theoretisch orientiert. Die praktische Umsetzung erfolgte dann durch J. Tinbergen in einem Rahmen, der das vorrangige Interesse empirischer Ökonomen darstellte: der Analyse von Konjunkturzyklen.¹⁷⁵ Anders als die in Abschnitt B 2 b angesprochenen Ansätze von Moore, Mitchell oder Persons begann Tinbergen jedoch nicht auf der Ebene empirischer Daten, anhand derer Regelmäßigkeiten entdeckt und Erklärungen gesucht wurden, sondern auf der Ebene theoretischer Modelle für eine (Makro-) Ökonomie.¹⁷⁶

Er ging dabei von dem *Propagation*-Ansatz von Frisch aus. Dieser hatte jedoch eher hypothetischen Charakter. Ein vollständiges Modell für eine Volkswirtschaft, dessen Mechanismus Zyklen induzierte, existierte bislang nicht. Es gab zwar eine Reihe von Theorien, doch waren diese noch keiner systematischen statistischen Überprüfung unterzogen worden.¹⁷⁷

data in deciding among competing hypotheses. Koopmans would agree with Friedman, that the significance levels of such tests were unknown [siehe nächster Abschnitt, T. R.]. But he retained the faith of Tinbergen and Frisch that their methods would surely lead to the discovery of a true macrodynamic structure." Epstein (1987, S. 55). Die Situation, daß man nur über wenige, nicht beeinflussbare Daten verfügen konnte, für die sich mehrere Modelle anpassen ließen und zwischen denen nicht anhand eines formalen Kriteriums entschieden werden konnte, führte Koopmans zu der Überzeugung, daß die ökonomische Theorie A-priori-Informationen bereitstellen mußte, um die Zahl der möglichen Hypothesen von vornherein zu reduzieren. Technisch bedeutete dies vor allem die Einführung von Nullrestriktionen in Strukturgleichungssystemen. Vgl. ebda., S. 53.

¹⁷⁴ Nach Morgan (1990, S. 240) war seine Arbeit "theoretical, highly technical in nature and probably understood by relatively few econometricians." Qin (1993, S. 18) sieht dagegen den Grund darin, daß seine "acceptance of probability theory appeared fragmentary and technical, i. e. not related to any of economists' reasoning."

¹⁷⁵ Jan Tinbergen [1903-1994] promovierte nach einem Studium in Leiden 1929 mit der Arbeit *Minimumproblemen in de natuurkunde en de ekonomie* in Physik und Ökonomie bei Paul Ehrenfest, war von 1933 bis 1973 Professor für Ökonomie an der Universität Leiden, arbeitete bis 1945 für das Statistische Zentralbüro und bis 1955 für das Niederländische Zentrale Planungsbüro, dessen Gründungsdirektor er war.

¹⁷⁶ Die Denkweise war in diesem Zusammenhang noch stark von der Physik geprägt. Siehe dazu den aufschlußreichen Beitrag von Boumans (1993). Andererseits entwickelte Tinbergen aber auch einen nicht unerheblichen Pragmatismus, den seine Kenntnis wirtschaftsstatistischer Daten durch seine Arbeit für das Statistische Amt mit sich brachte.

Ein erstes umfassendes Modell konstruierte Tinbergen für die holländische Wirtschaft von 1923 bis 1935.¹⁷⁸ Dieses Modell enthielt 31 Variablen und neben 6 definitorischen Identitäten 16 Gleichungen, die in der Regel ein oder zwei erklärende Variablen enthielten. Die Konstruktion eines solchen, für die Zeit ungewöhnlich komplexen Modells war mit einer Reihe von pragmatischen Entscheidungen verbunden, insbesondere was die Auswahl der relevanten Variablen anbetraf:

“One cannot know *a priori* what variables are necessary and what can be neglected in the explanation of the central phenomena that are under consideration. It is only during the actual work, and especially after the statistical verification of the hypotheses, that this can be discovered.”¹⁷⁹

Die Aufstellung des endgültigen Modells erfolgte in einem iterativen Prozeß: Theoretische Überlegungen bildeten den Ausgangspunkt, durch die Inspektion der vorliegenden Daten wurden diese Überlegungen revidiert und anschließend geprüft, ob die revidierten Modelle ökonomisch sinnvoll erschienen. Ein zusätzliches Kriterium (dies war die “statistical verification”) bot die visuelle Inspektion der Übereinstimmung der abhängigen Variablen (Zeitreihe) y_t mit der angepaßten Reihe sowie der Residualreihe.

Mit dieser Modellform waren noch keine Aussagen über die “Zyklizität” der Ökonomie verbunden. Um solche Aussagen treffen zu können, mußte aus den 22 Gleichungen eine “finale” Gleichung bestimmt werden, die nur noch eine Variable enthielt und den Zeitpfad des Systems wiedergeben sollte.¹⁸⁰ Diese finale Gleichung lautete

$$(1) \quad Z_t = 0.15Z_{t-1} + 0.26Z_{t-2} - 34.7.$$

Um zu überprüfen, ob die Parameterwerte eine gedämpfte Schwingung erzeugen, ermittelte Tinbergen die Wurzeln der charakteristischen Gleichung von (1) und fand als Lösung komplexe Zahlen. Er resümierte: “it is a well known fact that [...] in such a case [we have a Z] function showing a damped cyclic movement.”¹⁸¹

Betrachtet man hingegen Gleichung (1) genauer, so sieht man, daß diese Parameterwerte *keinen* Zyklus implizieren. Tinbergen hatte sich hier verrechnet; die Wirtschaft wies nach seiner finalen Gleichung keine zyklische Entwicklung auf.¹⁸²

¹⁷⁷ Das war die Schlußfolgerung eines Überblicks, den er 1935 in der *Econometrica* veröffentlichte und der auf seiner Arbeit für das Statistische Zentralbüro basierte. Vgl. Tinbergen (1935). Siehe zum folgenden auch ausführlich Morgan (1990, S. 101ff).

¹⁷⁸ Vgl. Tinbergen (1937).

¹⁷⁹ Ebda., S. 9, zitiert nach Morgan (1990, S. 103). Hervorhebung im Original.

¹⁸⁰ [...] the final equation represented the structure of the Dutch economy because the coefficients of the difference terms depended on the estimated coefficients of the system of 22 elementary equations which formed the model. So, the equation was not itself a regression equation, but it depended on all those that were.” Morgan (1990, S. 106). Zur Berechnungsweise siehe Tinbergen (1937, S. 55) sowie Wallis (1977).

¹⁸¹ Tinbergen (1937, S. 57), zitiert nach Neuberg (1995, S. 374).

¹⁸² Diese Tatsache war Morgan (1990, S. 106f) entgangen. Die Berechnung der Lösungen der “charakteristischen Gleichung” ist eine allgemeine Lösungsmethode für Differenzgleichungen beliebiger Ordnung. Für den Spezialfall von Differenzgleichungen zweiter Ordnung ist die Berechnung erheblich einfacher: Zyklische Schwingungen ergeben sich nur

Im selben Jahr wie Schumpeters *Business Cycles* erschien auch Tinbergens zweite Konjunkturzyklenuntersuchung, deren erster Band schon im Titel das statistische *Testen von Theorien* zum Programm erhob.¹⁸³ Diese von den Vereinten Nationen in Auftrag gegebene Studie, die die kurz vorher ebenfalls im Auftrag der Vereinten Nationen von G. Haberler zusammengestellten Konjunkturtheorien quantifizieren sollte, schloß unmittelbar an seine Analyse der holländischen Wirtschaft an. Nun konnte er jedoch bereits die Arbeit von Koopmans rezipieren. Tinbergen begann zunächst mit einer Erläuterung der Ansätze von Frisch und Koopmans und betonte die Bedeutung einer multiplen Analyse. Zwei Größen wie etwa der Preis und die Produktion eines Gutes seien in der Regel nicht direkt miteinander kausal verbunden, sondern in einer mehr oder weniger komplizierten Kette von Beziehungen. Diese gelte es in der Analyse zu "entdecken".¹⁸⁴

Dieser methodischen Einleitung folgten drei Fallstudien über Investitionsverhalten, die die Methoden illustrieren sollten. Wiederum stützte Tinbergen sich bei seinem iterativen Vorgehen auf die bereits erwähnten graphischen Darstellungen, um die endgültigen Beziehungen zu erhalten.¹⁸⁵

Anschließend wurden die Modelle getestet. Tinbergen überprüfte, ob sich die ermittelten Koeffizienten¹⁸⁶ in Aufstiegs- und Abstiegsphasen der Zyklen sowie in verschiedenen Unterperioden unterschieden und ob die Residuen normalverteilt und seriell unkorreliert waren. Schließlich untersuchte er die Modelle anhand von Frischs "bunch maps" auf Kollinearität und ermittelte für die Koeffizienten Fehlergrenzen nach der von Koopmans vorgeschlagenen Methode.

Der zweite Band versuchte nun, die im ersten Band dargelegte und anhand von Beispielen illustrierte Methode im größeren Rahmen anzuwenden und damit den Auftrag der Vereinten Nationen einzulösen. Als Datengrundlage wählte Tinbergen die US-amerikanische Ökonomie von 1919 bis 1932. Um die vorliegenden verbalen Theorien einer quantitativen Analyse zugänglich zu machen, war es zunächst nötig, diese in formale Modelle umzuformulieren. Tinbergen entschloß sich, aus den vorhandenen Theorien eine Kombination zu bilden, die er dann in ein formales Modell übersetzte und mit seinem bereits für die holländische Wirtschaft angewandten iterativen Verfahren revidierte. Das endgültige Modell bestand schließlich aus 71 Variablen und 48 Gleichungen. Im Gegensatz zum ersten Band wurde dieses Modell jedoch nicht mehr intensiven Tests ausgesetzt. Die ermittelten Gleichungen wurden nicht auf andere Länder oder Zeitabschnitte übertragen und waren somit lediglich eine Anpassung für die USA für 14 Jahre.¹⁸⁷ Wie auch schon im

für Parameterwerte ϕ_1 und ϕ_2 , die die Ungleichung $\phi_2 < -1/4 \phi_1$ erfüllen. Siehe dazu Rahlf (1996b).

¹⁸³ Vgl. Tinbergen (1939a). Wir wollen hier nur auf den methodischen Aspekt eingehen und die wirtschaftstheoretischen und -politischen Gesichtspunkte außer acht lassen.

¹⁸⁴ Vgl. Tinbergen (1939a, S. 12).

¹⁸⁵ Die Trendfrage spielte hier allerdings keine eigenständige Rolle mehr. Fast alle Variablen gingen nicht in ihren Originalwerten, sondern als Abweichungen von einem gleitenden Mittelwert in die Berechnungen ein. Vgl. ebda., S. 47.

¹⁸⁶ Zur Bestimmung der Koeffizienten bediente er sich allerdings nicht der von Koopmans in Anlehnung an Fisher vorgeschlagenen Maximum-Likelihood-Methode, sondern der gewöhnlichen Kleinstquadratrechnung.

holländischen Modell konstruierte er schließlich eine "finale Gleichung", um zu überprüfen, ob das System als Ganzes ein zyklisches Muster aufwies.¹⁸⁸ Tinbergens erster Band wurde kurz nach seinem Erscheinen ausführlich von J. M. Keynes besprochen.¹⁸⁹ Diese mittlerweile berühmte Kritik ist zahlreichen, auch kontroversen Interpretationen unterzogen worden. Während z. B. Morgan Tinbergen in nahezu allen Punkten verteidigt und die Kritik von Keynes als unzutreffend bezeichnet, ist McAleer der Ansicht, daß Keynes die Literatur über unit-roots, Kointegration, Tests auf Strukturwandel und Ausreißer sowie einige andere "moderne" Aspekte vorweggenommen hat.¹⁹⁰ Einig ist man sich lediglich darüber, daß Keynes den Band nur oberflächlich gelesen und mehrere technische Einzelheiten nicht verstanden hatte. So war ihm nicht klar, wie eine Linearkombination Zyklen erzeugen könne; weiterhin ging er z. B. davon aus, daß eine Trendgerade dadurch bestimmt werde, daß man den ersten und den letzten Datenpunkt miteinander verbinde.¹⁹¹

¹⁸⁷ Später bemerkte er hinsichtlich der Tests: "I was not primarily interested in these things and I have to admit that at present they are done much more seriously than they were at that time." Tinbergen in Magnus/Morgan (1987, S. 128). Die Bedeutung von *t*-Werten war ihm zu diesem Zeitpunkt schlichtweg unbekannt. Siehe ebda., S. 126.

¹⁸⁸ Zur konkreten Berechnungsform siehe Tinbergen (1939b, S. 130ff). Nicht nur dafür war die genaue Ermittlung der Koeffizienten und nicht nur ihre Signifikanz oder ein "richtiges" Vorzeichen wichtig: "The coefficients [of a dynamic system represent] the structure of society ... [where] the coefficients may also be changed as a consequence of policy, and the problem of finding the best stabilizing policy would consist in finding such values for the coefficients as would damp down the movements [of the endogenous variables] as much as possible. *The outstanding importance of the numerical values of the coefficients may be clear from these few considerations.*" Tinbergen (1939b, S. 18), zitiert nach Epstein (1987, S. 48). Hervorhebungen im Original.

¹⁸⁹ Vgl. Keynes (1939), Keynes (1940). Lord [seit 1942] John Maynard Keynes [1883-1946] ist vielleicht der bedeutendste, neben Karl Marx aber sicher der einflußreichste Ökonom des 20. Jahrhunderts gewesen. Er war der Sohn von J. N. Keynes, der das Standardlehrbuch *Scope and Method of Political Economy* verfaßt hatte. J. M. Keynes studierte nach seinem Schulbesuch in Eton 1902 bis 1905 am *King's College* in Cambridge Mathematik. 1906 absolvierte er das *Civil Service*-Examen und begann als *Junior clerk* am *India Office* in London. 1908 wurde er *lecturer* für Ökonomie an der *Cambridge University* – ohne einen ökonomischen Abschluß zu besitzen oder eine ökonomische Publikation vorweisen zu können. 1911 wurde er Herausgeber des schon damals bedeutenden *Economic Journal*, vier Jahre später Mitarbeiter des Schatzamtes. Dort stieg er schnell in leitende Positionen auf und nahm 1919 als Vertreter des Schatzkanzlers an den Friedensverhandlungen in Versailles teil. Anschließend übte er in London und Cambridge zahlreiche publizistische und beratende Tätigkeiten aus und publizierte nach einer Vielzahl wissenschaftlicher Beiträge 1936 schließlich sein Hauptwerk, die *General Theory of Employment, Interest and Money*. Eine seiner ersten größeren wissenschaftlichen Arbeiten war das 1921 erschienene *Treatise on Probability*, das in einer früheren Fassung von 1909 bereits der Anlaß für seine Wahl zum *fellow* des *King's College* war. 1943 kam Keynes der Bitte von A. Cowles nach, Präsident der *Econometric Society* zu werden und übte dieses Amt 1944 und 1945 aus.

¹⁹⁰ Vgl. Morgan (1990, S. 121ff), McAleer (1995, S. 105): "Keynes outlined a research program, perhaps unconsciously, which subsequently led to the development of numerous econometric techniques that are now widely used in applied econometrics." Siehe zur Diskussion auch Epstein (1987, S. 50ff), Hendry/Morgan (1995b, S. 52-60), Lawson (1989), Pesaran/Smith (1985), Conniffe (1992) sowie Stone (1978).

¹⁹¹ Vgl. Tinbergen in Magnus/Morgan (1987, S. 129). Die Kontroverse fand unmittelbar nach einer lebensbedrohlichen, ihn bis zu seinem Tod einschränkenden Krankheit und direkt vor

Dennoch ist ein Großteil der von Keynes geäußerten Kritik berechtigt. So hinterfragte er die von Tinbergen durchgeführte Formalisierung und Konkretisierung der Konjunkturtheorien, insbesondere die Wahl der "time-lags", deren Ausdehnung in den verbalen Theorien nur vage formuliert worden sei:

"To the best of my understanding, Prof. Tinbergen is not presented with his time-lags, as he is with his qualitative analysis, by his economist friends, but invents them for himself. This he seems to do by some sort of trial-and-error method. That is to say, he fidgets about until he finds a time-lag which does not fit in too badly with the theory he is testing and with the general presuppositions of his method."¹⁹²

Auch die Wahl des Trends und die Analyse anhand von *Trendabweichungen* sei willkürlich, die grundlegenden Variablen müßten statt dessen in ihren Niveaus zueinander in Beziehung gesetzt werden: "[...] should not the trends of the basic factors be allowed to be reflected in a trend of the resulting phenomenon?"¹⁹³ Weiterhin kritisierte Keynes Tinbergens Annahme einseitiger Kausalitäten:

"What happens if the phenomenon under investigation itself reacts on the factors by which we are explaining it? For example, when he investigates the fluctuations of investment, Prof. Tinbergen makes them depend on the fluctuations of profit. But what happens if the fluctuations of profit partly depend (as indeed, they clearly do) on the fluctuations of investment?"¹⁹⁴

Eine Reihe weiterer Kritikpunkte war dagegen unzutreffend. Keynes warf Tinbergen vor, viele Annahmen nicht überprüft zu haben, die Tinbergen jedoch explizit überprüft hatte, so z. B. die Frage der zeitlichen Stabilität der errechneten Koeffizienten. Seine Forderung, Teilperioden zu vergleichen, war von Tinbergen erfüllt worden.¹⁹⁵ Bedenkt man jedoch die Vielzahl der Ad-hoc-Entscheidungen und Korrekturen, die Tinbergen vornahm sowie die schmale Datengrundlage, so ist Keynes Gesamteindruck, daß es sich bei der Arbeit vor allem um "[...] no more than a piece of historical curve-fitting and description"¹⁹⁶ handelte, tendentiell zuzustimmen. Keynes ging prinzipiell davon aus, daß wahrscheinlichkeitstheoretische Kalküle lediglich für Situationen angemessen seien, die Glücksspielcharak-

seinem zweiten Engagement für die britische Regierung statt. Nach Beginn des Krieges wurde Keynes *Advisor to the Chancellor of the Exchequer*, Direktor der *Bank of England*, schließlich Verhandlungsführer bei Verhandlungen zur Kriegs- und Nachkriegsfinanzierung. Vielleicht erklären diese Ereignisse neben der inhaltlichen Ablehnung die begrenzte Aufmerksamkeit, die Keynes der ökonometrischen Entwicklung schenkte.

¹⁹² Keynes (1939, S. 565).

¹⁹³ Ebda. Für die Beziehung zwischen Investitionen, Profiten und Zinsen fragte er dagegen: "[...] some economists would argue that it is the *difference* between these two factors which matters, rather than their absolute amounts." Hervorhebung im Original. Auch hier hatte Tinbergen Abweichungen von einem Trend zugrundegelegt.

¹⁹⁴ Keynes (1939, S. 561).

¹⁹⁵ An anderer Stelle sprach sich Keynes noch deutlicher aus: "The coefficients arrived at [in dynamischen Modellen, T. R.] are apparently assumed to be constant for ten years or for a longer period. Yet, surely we know they are not constant. *There is no reason at all, why they should not be different every year.*" Zitiert nach Lawson (1989, S. 247). Hervorhebung von uns.

¹⁹⁶ Keynes (1939, S. 566).

ter hätten. Die ökonomische Realität war seiner Meinung nach zu komplex, als daß man sie durch statistische/stochastische Gesetze erfassen könne. Bereits in seiner Abhandlung über den Begriff der Wahrscheinlichkeit hatte er sich skeptisch gegenüber der Möglichkeit geäußert, die Wahrscheinlichkeitstheorie auf Gebiete außerhalb der Biologie und Physik zu übertragen.¹⁹⁷

In seiner Antwort auf Keynes' Kritik konnte Tinbergen zwar eine Reihe von Mißverständnissen korrigieren,¹⁹⁸ das Grundproblem, das Keynes in seinem Kommentar auf Tinbergens Antwort auf den Punkt brachte, war aber nicht aus der Welt zu schaffen:

"It will be remembered that the seventy translators of the Septuagint were shut up in seventy separate rooms with the Hebrew text and brought out with them, when they emerged, seventy identical translations. Would the same miracle be vouchsafed if seventy multiple correlators were shut up with the same statistical material?"¹⁹⁹

Die Kontroverse zog eine Reihe weiterer Diskussionen nach sich. Ähnliche Vorbehalte wie Keynes äußerte M. Friedman. Als exploratives Verfahren zur Ermittlung von "tentative hypotheses" sei das Vorgehen angemessen, nicht jedoch als Test von Theorien:

"Tinbergen's results cannot be judged by ordinary tests of statistical significance. The reason is that the variables with which he winds up [...] have been selected after an extensive process of trial and error *because* they yield high coefficients of correlation [...]. They are tautological reformulations of *selected* economic data."²⁰⁰

Von einer anderen Warte aus kritisierte Frisch Tinbergen.²⁰¹ Frisch beanstandete wie schon Keynes die Willkür in der Wahl der "lags", widmete sich aber hauptsächlich dem Problem der "Autonomie" und führte den Begriff des "strukturellen Systems" ein. In einem strukturellen System sind die Beziehungen innerhalb einer Gleichung invariant gegenüber Veränderungen von anderen Beziehungen innerhalb des Systems, die Gleichungen sind "irreduzibel", autonom oder in Frischs Diktion "coflux". Tinbergen hatte seiner Ansicht nach jedoch keine strukturellen Systeme konstruiert, sondern Gleichungen, die nur einen geringen Grad an Autonomie aufwiesen. Insofern waren sie für die von Tinbergen anvisierte politische Steuerung ungeeignet.²⁰²

Trotz aller Kritik an Tinbergens Arbeit darf nicht übersehen werden, daß sie einen bleibenden Einfluß auf die Entwicklung der Ökonometrie ausübte und die Mög-

¹⁹⁷ Diese Ansicht vertrat er bereits in Keynes (1921).

¹⁹⁸ Vgl. Tinbergen (1940a).

¹⁹⁹ Keynes (1939, S. 155f). Im gleichen Sinne auch Rostow (zitiert nach Klump (1993, S. 178): "Men observing honestly the same set of data emerged with quite different explanations. Each explanation depended directly on the theoretical presuppositions."

²⁰⁰ Friedman (1940, S. 659). Hervorhebung im Original.

²⁰¹ Vgl. Frisch (1938 [1995]). Dieses Memorandum, das Frisch für eine Konferenz vorbereitet hatte, blieb unveröffentlicht, zirkulierte jedoch unter verschiedenen Ökonometrikern und hatte großen Einfluß auf die spätere Arbeit der *Cowles Commission*.

²⁰² Vgl. auch Wallis (1994a, S. xi).

lichkeit, eine gesamte Ökonomie anhand eines Gleichungssystems zu charakterisieren, zumindest innerhalb ökonometrischer Kreise seitdem als sinnvoll und notwendig erschienen. Diesen grundsätzlichen Aspekt, der im übrigen auch eines der Hauptziele des Programms der *Econometric Society* darstellte, wurde in einer Verteidigung von Tinbergens Ansatz auch noch einmal von T. Haavelmo und T. Koopmans hervorgehoben.²⁰³

Frischs Ansatz, auf den wir hier nicht in den technischen Einzelheiten eingehen wollen,²⁰⁴ legte den Grundstein für eine langanhaltende Diskussion in der Ökonometrie bezüglich der Identifikation und Schätzung von strukturellen Gleichungssystemen. Die Grundlagen für diese Diskussion wurden im folgenden vor allem von T. Haavelmo gelegt, dem wir uns im folgenden Abschnitt zuwenden wollen.

d. Übertragung der "klassischen" Sicht in die Ökonometrie

Der Grundstein der heute dominierenden inferenzstatistischen Sicht in der Ökonometrie wurde ohne Zweifel 1944 durch T. Haavelmo gelegt.²⁰⁵ Haavelmo, dem 1989 "für seine Klärung der wahrscheinlichkeitstheoretischen Grundlagen ökonometrischer Methodik und seine Analyse simultaner ökonomischer Strukturen"²⁰⁶ der Nobelpreis für Wirtschaftswissenschaften verliehen wurde, hatte in Anlehnung an Frischs Memorandum von 1938 darauf hingewiesen, daß in einem vollständigen System, wie es von Tinbergen konstruiert wurde, bei Vorliegen simultaner Beziehungen die Kleinstquadratschätzer verzerrt waren. Prinzipiell war das Problem bereits seit E. Workings Arbeit von 1927 über Angebots- und Nachfragefunktionen bekannt. Wenn zwei Variablen in einem System miteinander verbunden waren,

$$(2) \quad Y = \alpha X + \varepsilon_1 \quad (\text{Nachfrage})$$

$$Y = \beta Y + \varepsilon_2 \quad (\text{Angebot})$$

konnten, auch wenn ε_1 und ε_2 unabhängig waren, die Parameter mit dem herkömmlichen Kleinstquadratansatz isoliert nicht unverzerrt geschätzt werden.²⁰⁷ Haavelmos Lösung bestand in einer Maximum-Likelihood-Schätzung des gesamten Systems. Auch hier wollen wir nicht auf die technischen Einzelheiten eingehen, sondern uns der dem Haavelmo-Ansatz zugrundeliegenden Inferenzkonzeption

²⁰³ Siehe Haavelmo (1943a), Koopmans (1941). Auch Tinbergen hat dies noch einmal betont, vgl. Tinbergen (1940b).

²⁰⁴ Siehe dazu Hendry/Morgan (1995b, S. 57-60) sowie Boumans (1995).

²⁰⁵ Trygve Haavelmo [1911 -] war ein Schüler von R. Frisch. Von 1932 bis 1942 sowie von 1943 bis 1946 hielt er sich in den USA auf, unter anderem bei der *Cowles Commission*. 1947 wurde er Professor für Ökonomie in Oslo. Seine 1944 in der *Econometrica* als Supplement veröffentlichte Arbeit zirkulierte bereits ab 1941 in vervielfältigter Form. Die Arbeit bezog sich zwar nicht explizit auf Keynes' Kritik an Tinbergen, kann aber doch in weiten Teilen als eine Antwort darauf verstanden werden. Vgl. Lawson (1989, S. 248). Der Beitrag von 1943 (Haavelmo (1943b)) ist ein Ausschnitt aus der 1944 veröffentlichten Arbeit.

²⁰⁶ Offizielle Preisbegründung, abgedruckt in Grüske (1994, S. 63).

²⁰⁷ Vgl. Haavelmo (1943b, S. 2).

zuwenden, denn in dieser "probability-revolution" wird die eigentliche Bedeutung des Beitrags gesehen.²⁰⁸

Wie wir im vorigen Abschnitt gesehen hatten, ging Koopmans davon aus, daß die beobachteten bzw. vorliegenden ökonomischen Daten eine (hypothetische) Zufallsrealisation aus einer (hypothetischen) infiniten Population darstellten. Er bezog sich dabei explizit auf R. A. Fisher. Haavelmo ging ebenfalls von einer solchen stochastischen Grundannahme aus. Seiner Ansicht nach war zwar in der ökonomischen Forschung eine Reihe von statistischen Methoden verwendet worden, doch habe man dabei nicht die stochastischen Grundlagen, die damit verbunden seien, beachtet:

"[...] it has been considered legitimate to use some of the tools developed in statistical theory *without* accepting the very *foundation* upon which statistical theory is built. For *no tool in the theory of statistics has any meaning* – except, perhaps for descriptive purposes – *without being referred to some stochastic scheme*."²⁰⁹

Im folgenden widmete er sich der Bedeutung von Modellen und Gesetzen in der Ökonomie und Statistik, der Bedeutung von Wahrscheinlichkeitsaussagen für ökonomische Daten, dem Problem des Hypothesentestens, der Schätzung von Parametern und schließlich der Prognose. Haavelmo ging davon aus, daß es für den Erkenntnisprozeß 1. notwendig sei, theoretische Modelle zu bilden und 2. diese Modelle an der Realität zu prüfen. Die Erfahrung zeige nun, daß sich die Realität niemals in exakter Übereinstimmung mit einem Modell befinde, insofern also jedes Modell falsch sei.²¹⁰ Dennoch seien solche "falschen" Modelle von Nutzen:

"Therefore we shall not only have to be satisfied with broader statements than the ones usually implied by an 'exact' model, but we shall also have to adopt a particular kind of model, namely such models as permit statements that are not *implications*, but merely have a certain chance of being true. This will lead us to a probabilistic formulation of theories that are meant to be applied."²¹¹

Zunächst war für ihn die Unterscheidung verschiedener Ebenen von zentraler Bedeutung. Es sei prinzipiell nicht möglich, genau die Größen zu messen, die in einer Theorie formuliert werden, wodurch es zu "Diskrepanzen" zwischen Theorie

²⁰⁸ Diese Interpretation wird vor allem von Morgan (1987, 1990) vertreten. Siehe jedoch bereits Menges (1959b) sowie Mirowski (1989c), Lawson (1989) und Heckman (1992). Zu den von Haavelmo angesprochenen technischen Problemen der Schätzung von simultanen Strukturgleichungssystemen siehe Epstein (1987, S. 56ff) sowie Hendry/Morgan (1995b, S. 60-66).

²⁰⁹ Haavelmo (1944, S. iii). Hervorhebungen im Original. Andererseits betonte er jedoch, daß diese stochastische Erklärungskonzeption nur eine neben anderen möglichen sei. Gemeinsam sei jedoch allen, daß sie "Erfindungen" darstellten, keine verborgenen Wahrheiten, die es zu entdecken gäbe: "At least this is one type of 'explanation'. Other types may be chosen. But whatever be the 'explanations' we prefer, it is not to be forgotten that they are all our own artificial inventions in a search for an understanding of real life; they are not hidden truths to be 'discovered.'" Ebda., S. 3.

²¹⁰ "In other words, such exact models are simply false in relation to the facts considered." Ebda., S. 1.

²¹¹ Ebda., S. 2. Hervorhebung im Original.

und Daten komme. Er unterschied "Beobachtungs-", "wahre" und "theoretische" Variablen. Letztere charakterisierte er wie folgt:

"We may express the difference by saying that the 'true' variables (or time functions) represent our ideal as to accurate measurements of reality 'as it is in fact,' while the variables defined in a theory are the true measurements that we should make if reality were actually in accordance with our theoretical model."²¹²

Durch diese Definition ergab sich eine Analogie zu naturwissenschaftlichen Experimenten. Andererseits räumte Haavelmo jedoch ein, daß der Ökonom sich lediglich in der Rolle des passiven Beobachters befinde und keine Experimente durchführen könne. Wenn die Größen mathematischer Gleichungen mit ökonomischen Variablen bezeichnet werden, könne man dies als (virtuelles) Experiment auffassen. Falls man von einem stochastischen Ansatz ausgehe, in dem alle Variablen einer multivariaten Verteilung unterliegen, dann sei kein Wert (keine Realisation) als Möglichkeit ausgeschlossen, den einzelnen Werten würden lediglich unterschiedliche Gewichte bzw. Wahrscheinlichkeiten zugeordnet:

"[...] the restrictions imposed might not absolutely exclude any value of the quantities considered; it might merely give different weights (or probabilities) to the various sets of possible values of the variable quantities. [...] *according to experience it has very often been found fruitful to interpret* such weights as a measure of actual 'frequency of occurrence.' [...] If the total weight ascribed to all the possible value-systems is finite, we can then say that the practical meaning of a set of value-systems that has a weight almost equal to zero according to the model is a hypothesis saying that Nature has a way^[213] of selecting joint value-systems of the corresponding 'true' variables that makes it 'practically impossible' that a system of observed values should fall within such a set. For the purpose of testing the theory against some other alternative theories we might then agree to deem the hypothesis tested false whenever we observe a certain number of such 'almost impossible' value-systems. That is, at the risk of making an error, we should then prefer to adopt another hypothesis under which the observations made are not of the 'almost impossible' type."²¹⁴

Das bedeutete: Die ökonomische Theorie stellte das Modell (die multivariate Verteilung, für Haavelmo gleichbedeutend mit einem Strukturgleichungsmodell) bereit, und "die Natur" zog daraus eine Stichprobe; dies waren dann die konkret vorliegenden Daten. Solch eine Vorstellung müsse nicht notwendigerweise zutreffen, aber: "*It has been found fruitful* to introduce a special calculus for deriving such hypotheses. This is the calculus of probability."²¹⁵ Eine solche "Wahrscheinlichkeit" müsse nicht notwendigerweise existieren:

²¹² Ebda., S. 5.

²¹³ Diese Redewendung verwendete Haavelmo mehrfach.

²¹⁴ Haavelmo (1944, S. 8f). Konkret ist darunter vermutlich eine Situation wie die folgende zu verstehen: In einer bivariaten Normalverteilung mit dem Erwartungswert $E[\mu_1, \mu_2]$ (die "wahren" Werte) ist ein konkretes Wertepaar (x_1, x_2) , das sehr weit von diesem bivariaten Erwartungswert entfernt ist, sehr unwahrscheinlich (die Natur hat einen Weg, der diesen Wert praktisch unmöglich macht). Wenn eine solche Beobachtung gemacht wird, kann man, mit dem Risiko, einen Fehler zu machen, davon ausgehen, daß die "Theorie" $E[\mu_1, \mu_2]$ falsch ist.

²¹⁵ Ebda., S. 10. Hervorhebung von uns.

“The question is not whether probabilities exist or not, but whether – if we proceed *as if* they existed – we are able to make statements about real phenomena that are ‘correct for practical purposes.’”²¹⁶

Die häufigkeitstheoretische Definition, die mit diesem Ansatz verbunden ist, erschien ihm daher unpassend:

“Above we considered ‘frequency of occurrence’ as a practical counterpart to probability. But in many cases such an interpretation would seem rather artificial, e.g., for economic time series where a repetition of the ‘experiment,’ in the usual sense, is not possible or feasible. Here we might then, alternatively, interpret ‘probability’ simply as a measure of our *a priori confidence* in the occurrence of a certain event.”²¹⁷

Im Hinblick auf die Möglichkeit einer Verallgemeinerung der in konkreten Daten gefundenen Struktur über die beobachteten Daten hinaus (in Haavelmos Motivation vorrangig die Prognose) betonte er: “We cannot give any *a priori* reason for such a supposition. We can only say, that *according to a vast record of actual experiences, it seems to have been fruitful to believe in the possibility* of such empirical inductions.”²¹⁸ Über den Vorgang der Konstruktion eines Modells könne dabei jedoch keine Aussage gemacht werden: “It is a creative process, an art” und beruhe auf “believe” und “experience”.²¹⁹ Hauptaufgabe der Ökonometrie sei es, einfache Beziehungen zwischen abhängigen und einer geringen Anzahl unabhängiger Variablen zu bestimmen. Auch hier vertrat Haavelmo wiederum einen instrumentalistischen Standpunkt: “Whether or not such simple relations can be established must be decided by actual trials. *A priori* it can neither be shown to be possible nor proved impossible.”²²⁰ Sei jedoch ein Modell zustande gekommen, so folge die praktische Umsetzung der stochastischen Konzeption: das Testen von Hypothesen. In diesem Bereich habe erst vor kurzem die Theorie von J. Neyman und E. S. Pearson eine Grundlage geschaffen, die die bis dahin vorhandene Konfusion in der theoretischen Statistik beseitigt habe.²²¹ Nach dem oben genannten Schema bestehe der Test dann darin, das theoretische Modell (die multidimensionale Verteilung) mit den vorliegenden Daten zu vergleichen, indem man die Wahrscheinlichkeit dieser konkreten Realisation unter Annahme der Richtigkeit des Modells betrachte:

²¹⁶ Ebda., S. 43. Hervorhebung im Original.

²¹⁷ Ebda., S. 48. Hervorhebung im Original.

²¹⁸ Ebda., S. 10. Hervorhebung von uns.

²¹⁹ Ebda. Ebenso S. 29: “The construction of systems of autonomous relations is, therefore, a matter of intuition and factual knowledge; it is an art.”

²²⁰ Ebda., S. 23. Siehe dazu auch seine Ausführungen in Abschnitt 17: “The meaning of the phrase ‘To formulate theories by looking at the data’”, die ohne Zweifel eine Verteidigung der Vorgehensweise von Tinbergen darstellen.

²²¹ Vgl. ebda., S. 60. Auf die Darstellung der Neyman-Pearson-Testtheorie brauchen wir hier nicht noch einmal einzugehen. Siehe dazu oben, Abschnitt A 3 e.

“For the purpose of testing hypotheses it is not even necessary to assume that the sample could actually be repeated. We make hypothetical statements *before we draw the sample*, and we are only concerned with whether the sample rejects or does not reject an a priori hypothesis.”²²²

Das Modell selbst, das “wahre” Wahrscheinlichkeitsgesetz, bleibe dabei unbekannt und müsse als Axiom eingeführt werden.²²³ Diesem allgemeinen Ansatz des Modelltestens folgten dann Ausführungen über die Schätzung eines Strukturgleichungssystems²²⁴ und abschließende Bemerkungen über die Prognose anhand der geschätzten Modelle.

Die Bedeutung von Haavelmos Beitrag für die weitere Entwicklung ist kaum zu überschätzen. Vor allem die einflußreiche *Cowles Commission* sah es als ihre Aufgabe an, den von Haavelmo vorgezeichneten Weg weiterzugehen, konkrete Schätztechniken zu entwickeln und damit makroökonomische Modelle zu evaluieren.²²⁵ Vor der sichtbaren Umsetzung dieses Programms kam es jedoch zu einer Auseinandersetzung um den Haavelmo-Ansatz, der berühmt geworden ist: die “Measurement-without-Theory”-Debatte zwischen T. C. Koopmans und R. Vining.²²⁶

Ausgangspunkt war eine Besprechung des 1946 erschienenen umfangreichen Werkes *Measuring Business Cycles* von A. Burns und W. C. Mitchell²²⁷ durch T. C. Koopmans. Koopmans begann seine Kritik bezeichnenderweise mit einer Analogie aus der “Himmelsmechanik”: dort habe es ein “Kepler-Stadium” und ein “Newton-Stadium” gegeben. Kepler habe sich wie schon vor ihm Tycho Brahe vorrangig um das Aufdecken empirischer Regularitäten in den Planetenbewegungen bemüht. Seine darauf aufbauenden Modelle hätten sich zum Teil als falsch,

²²² Haavelmo (1944, S. 70). Hervorhebung von uns.

²²³ Vgl. ebda., S. 49.

²²⁴ Eine “lesbare” Fassung dieses Kapitels erschien 1943, vgl. Haavelmo (1943b).

²²⁵ Schumpeter (1965, S. 1413) bemerkte etwa, daß Haavelmo “während seines Besuches in den Vereinigten Staaten, ohne ein Lehramt innezuhaben, einen Einfluß ausübte, der dem Lebenswerk eines Professors zum Ruhme gereichen könnte.” Die *Cowles Commission for Research in Economics* wurde 1932 von Alfred Cowles in Colorado Springs gegründet. Cowles, dessen ursprüngliche Intention die Förderung von Analysen zur Verbesserung von Aktienkursprognosen war und der deshalb die *Econometric Society* finanziell unterstützte (s. o.), rückte mehr und mehr von seinen ursprünglichen Interessen ab und entwickelte sich zu einem Mäzen, der die ökonometrische Forschung nun ganz allgemein unterstützte. Die Haupttätigkeit der Kommission bestand zunächst in der Veranstaltung von Tagungen und Seminaren, die von führenden Ökonomen und Statistikern, wie J. Schumpeter, J. Marschak oder R. A. Fisher, durchgeführt wurden. Darüber hinaus offerierte sie Flüchtlingen vor dem Nazi-Regime, wie etwa A. Wald, finanzielle Unterstützung. 1937 siedelte die Kommission, die nun bereits über einen Jahresetat von \$ 50.000 verfügte, nach Yale, 1939 nach Chicago um. Nachdem eine Reihe von Mitarbeitern kriegsbedingt ausgeschieden war, konnte Cowles 1943 J. Marschak als Forschungsdirektor gewinnen, der sich seiner neuen Aufgabe unmittelbar mit großem Engagement annahm. Er sammelte bis 1945 eine Reihe von herausragenden, mathematisch orientierten Wissenschaftlern um sich, darunter Tjalling Koopmans, Leonid Hurwicz, Herman Rubin, Lawrence Klein und T. W. Anderson, 1947 schließlich Kenneth Arrow. Damit waren drei spätere Nobelpreisträger zu dieser Zeit in der Kommission tätig. Vgl. Epstein (1987, S. 60ff, 75).

²²⁶ Koopmans (1947), Vining (1949a), Koopmans (1949), Vining (1949b).

²²⁷ Siehe oben, Abschnitt B 2 d.

zum Teil als irrelevant erwiesen. Der Theoretiker Newton konnte dagegen, zum Teil auf Keplers Beobachtungen basierend, die grundlegenden Gesetze der Gravitation entdecken und damit unter anderem auch Keplers Beobachtungen "erklären". Koopmans Ansicht nach ist die Intention von Burns und Mitchell mit derjenigen Keplers zu vergleichen. Ein solches Vorgehen sei zu begrüßen, wenn man davon ausgehen könne, daß andere Wissenschaftler auf ihrer Grundlage analog zu Newton Theorien entwickeln könnten. Für die Ökonomie erweise sich eine solche Reihenfolge jedoch nicht als vielversprechend. Hier müsse der Prozeß der Theoriebildung, der Beobachtung und der Messung miteinander verwoben sein. Burns und Mitchell beschränkten sich dagegen auf ein reines Messen und zusammenfassendes Beschreiben von Konjunkturzyklen:

"The toolkit of the theoretical economist is deliberately spurned. Not a single demand or supply schedule or other equation expressing the behavior of men of the technical laws of production is employed explicitly in the book [...]. The movements of economic variables are studied as if they were the eruptions of a mysterious volcano whose boiling caldron can never be penetrated."²²⁸

Die von Burns und Mitchell vorgenommene umfangreiche Charakterisierung von Zyklen stellte für Koopmans keinen Wert an sich dar, wenn sie nicht mit einer Erklärung und Prognose verbunden werde, denn ein wichtiges Kriterium wissenschaftlicher Analyse sei der soziale Nutzen durch wirtschaftspolitische Beratung:

"The criterion of social usefulness of scientific analysis gives us the right to discuss the merits of any particular approach to the problem of economic fluctuation on the basis of the guidance it gives to economic policy, even if such guidance were not claimed by the authors."²²⁹

Eine solche Analyse könne nur erfolgreich sein, wenn Theorien existierten, die die Beobachtungen erklären. In der Ökonomie seien solche Theorien nicht nur vorhanden, sondern darüber hinaus auch in statistische Modelle übertragbar:

"[...] according to that theory the relevant economic variables are determined by the simultaneous validity of an equal number of 'structural' equations (of behavior, of law or rule, of technology)."²³⁰

Wenn man die strukturellen Gleichungen kenne, könne dies darüber hinaus zu unserem Verständnis von Phänomenen in anderen Ländern und zu anderen Zeiten beitragen, "in the same way [...] in which the law of gravitation explains celestial and terrestrial phenomena alike."²³¹ Der seiner Ansicht nach interessanteste Aspekt der Arbeit betrifft die von den Autoren gefundene starke "random variation", die eine genauere Charakterisierung der Zyklen ihrer Ansicht nach erschwere. Genau dieser Punkt sei jedoch wiederum nur unter Zuhilfenahme der Theorie zu erklären, die in jüngster Zeit in Form von vollständigen Systemen von Strukturgleichungen,

²²⁸ Koopmans (1947, S. 163, 167).

²²⁹ Ebda., S. 166.

²³⁰ Ebda.

²³¹ Ebda., S. 167.

speziell in Form von stochastischen Differenzengleichungen, formuliert worden sei, in denen der gemeinsame Effekt der unberücksichtigten Faktoren als zufällig angesehen werden könne.²³² In solchen Systemen könne es unter bestimmten Bedingungen zu unregelmäßigen zyklischen Entwicklungen kommen.²³³ Das Ziel einer statistischen Untersuchung müsse es daher sein, ein solches System zu schätzen und unter Zugrundelegung einer Wahrscheinlichkeitsverteilung zu testen. Obwohl Haavelmo in der Besprechung nicht erwähnt wurde, ist doch offensichtlich die von Koopmans vertretene Position eine Propagierung seines Ansatzes. Es mag daher überraschen, daß R. Vining in seiner Antwort auf Koopmans sich auf Haavelmo bezog. Hierin ist jedoch nicht der Versuch zu sehen, Haavelmo für seine Argumentation zu vereinnahmen, wie dies Morgan behauptet.²³⁴ Im Gegensatz zu Koopmans legte er das Gewicht nicht auf die zweite, dritte und vierte Phase von Haavelmos Programm (Testen und Schätzen von Modellen sowie die Prognose), sondern auf die erste: die Konstruktion von vorläufigen theoretischen Modellen. Der entscheidende Punkt war für ihn, daß die theoretischen Modelle der Ökonomie (noch) nicht das soziale Pendant zur Newtonschen Mechanik darstellten.²³⁵

“When we think of the enormous body of factual knowledge digested and systematized in the other fields of variation and the meagerness of our own results from efforts to systematize, are we quite ready to leave Haavelmo’s first problem and launch into the last three problems in estimation theory?”²³⁶

Insofern sei das explorative Suchen von Hypothesen für die Ökonomie durchaus legitim. Hierfür hielt er die Neyman-Pearson-Maschinerie jedoch für ungeeignet: “Discovery has never been a field of activity in which elegance of conception and equipment is of prime consideration”.²³⁷ Seiner Ansicht nach mußte es keineswegs das erklärte Ziel der Wissenschaften sein, Inferenzaussagen zu treffen, die auf der “sampling theory” beruhen. Vining stützte sich hier vor allem auf die Argumentation, die Yule 1942 gegen die Überbetonung der “sampling theory” vorbrachte.²³⁸ Seine Wahl zwischen dem explorativen Vorgehen von Burns und Mitchell und der Postulierung von Strukturgleichungsmodellen von Koopmans fiel daher eindeutig aus:

²³² Vgl. ebda., S. 169.

²³³ Das ist die von Frisch in die Ökonomie eingeführte Grundidee der “random shocks”, die hier in ein Gleichungssystem eingebettet wurde.

²³⁴ Vgl. Morgan (1990, S. 253).

²³⁵ Vgl. Vining (1949a, S. 79).

²³⁶ Ebda., S. 83. Auf das in diesem Zusammenhang diskutierte Problem, ob das Verhalten von Individuen oder von aggregierten Größen (“aggregates” in Koopmans physikalischer Diktion, “populations” in Vinings biologischer Analogie) im Mittelpunkt stehen sollte, wollen wir hier nicht näher eingehen.

²³⁷ Vining (1949a, S. 78).

²³⁸ Wahrscheinlichkeitstheoretische Erwägungen lehnte Vining nicht prinzipiell ab, verwies aber auf die alternativen Ansätze der “kontinentalen Schule” (1949a, S. 85). Interessanterweise forderte er auch eine Wiederbelebung der alten Theorien der Häufigkeitsverteilungen. Damit war offensichtlich das auf K. Pearson zurückgehende (theorielose) System der Häufigkeitskurven gemeint. Siehe oben, Abschnitt A 3 c.

“In any event, I should feel much safer in bringing my little work before the Commissar of Research if that chair were occupied by Burns or Mitchell than if Koopmans were the occupant. I would feel that the first question put by Commissar Koopmans would be, ‘Where are your difference equations?’, and I would have to answer that I did not have any to speak of.”²³⁹

Koopmans wehrte ab. Die Analyse von Differenzgleichungen sei für ihn nur eine Notlösung:

“I am, for instance, quite unhappy about difference equations. They are clumsy instruments that treat time as if it comes in indivisible pieces of one year or one quarter each. The only excuses for their use are simplicity in exposition and the fact that statistical theory concerning the estimation of their parameters is further advanced than that for alternative, more realistic forms.”²⁴⁰

Auch hinsichtlich der Frage der Beziehung von Hypothesentests und explorativen Vorgehensweisen habe die statistische Theorie bislang keine überzeugende Antwort gefunden, insbesondere

“[...] between problems of ‘hypothesis-seeking’ and problems discussed in the theory of estimation or in the Neyman-Pearson theory of ‘hypothesis-testing’. This touches on unsolved problems at the very foundations of statistical theory, and I must confess that I do not see clearly through the issues involved. [...] For instance, we constantly take hints from the data regarding the choice of hypotheses to be tested from the same data – although we know that the degree of confidence to be placed in the test is affected by that practice.”²⁴¹

Dennoch hielt er prinzipiell an der Erfordernis, sich auf wenige Hypothesen zu beschränken und diese zu testen, fest. Die Notwendigkeit, Parameter zu schätzen und zu testen, sei bereits von Tinbergen hervorgehoben worden. Insbesondere wegen einer möglichen politischen Einflußnahme sei es nötig, die genauen Werte der Parameter zu kennen.²⁴²

Der letzte Punkt wurde schließlich von Vining noch einmal mit Bezug auf die Tinbergen-Keynes-Kontroverse aufgegriffen. Seiner Ansicht nach hätte die Mehrheit der Ökonomen den Argumenten von Keynes zugestimmt und Tinbergens Modell aus theoretischen Gründen als unbefriedigend empfunden. Tinbergens Analyse, ein “glorified multiple correlation project”²⁴³, habe beansprucht, ein – im algebraischen Sinne – vollständiges Gleichungssystem zu sein. Mittlerweile werde seiner Methode jedoch der Vorwurf gemacht, verzerrte Resultate zu erzielen, “verzerrt” im Sinne der theoretischen Statistik.²⁴⁴ Den Ökonomen interessierten jedoch ganz andere Probleme:

²³⁹ Vining (1949a, S. 85).

²⁴⁰ Koopmans (1949, S. 88).

²⁴¹ Ebda., S. 90.

²⁴² Vgl. ebda. Es ist allerdings kurios, daß er sich hier neben Tinbergen auch auf Friedman (1948) beruft, da Friedman in diesem Beitrag die Möglichkeit, strukturelle Modelle zur Analyse und Kontrolle von Konjunkturzyklen zu verwenden, ablehnte und der vehementeste Kritiker des strukturellen Ansatzes war. Friedman hatte, wie wir sahen, bereits unmittelbar nach dessen Erscheinen kritisch auf Tinbergens Ansatz reagiert, vgl. Friedman (1940).

²⁴³ Vining (1949b, S. 491).

²⁴⁴ Gemeint ist hier der oben erwähnte “Haavelmo-bias” der Kleinstquadratschätzung.

"[...] many economists, while commending the reduction in bias of estimation, would be more impressed with the possibilities of biases introduced by the philosophical postulates of the 'economic analysis'. Tinbergen computed standard errors, but as then computed they are not regarded as valid measures of reliability. Confidence intervals, it is contended, may now be computed appropriate for the model adopted. But the meaning of 'confidence' must be construed with care."²⁴⁵

Die Bedeutung beziehe sich nicht auf das Modell, obwohl dies das entscheidende Kriterium sei. Letztendlich müsse der Leser entscheiden, ob durch die Modelle von Tinbergen oder Koopmans der Ökonomie ein größerer Dienst erwiesen werde oder durch explorative, Fakten beschreibende und hypothesenbildende Arbeiten, wie diejenige von Burns und Mitchell oder andere des NBER.²⁴⁶

Die größte Schwäche in Koopmans Argumentation war sicher, daß er die Überlegenheit des Strukturgleichungsansatzes lediglich behauptete, aber noch keine erfolgreichen Beweise für dessen Überlegenheit vorlegen konnte. Diese Schwäche wurde im folgenden jedoch rasch behoben, und ungeachtet der Kritik²⁴⁷ von Vining entwickelte sich der Haavelmo-Ansatz schnell zur maßgeblichen Sichtweise.²⁴⁸

Wie bereits erwähnt wurde er insbesondere von der einflußreichen *Cowles Commission* zum Paradigma erhoben. Dessen Vorsitzender, J. Marschak, sah die

²⁴⁵ Vining (1949b, S. 94).

²⁴⁶ Vining erwähnt in diesem Zusammenhang die Arbeiten von Abramowitz (1950) und Friedman/Kuznets (1945).

²⁴⁷ Es muß allerdings betont werden, daß im Rahmen des NBER auch weiterhin (z. B. von M. Friedman oder S. S. Kuznets) zahlreiche bedeutende "nicht-ökonomische" Studien veröffentlicht wurden, die an die in Abschnitt B 2 d beschriebene Entwicklung anknüpften. Hier wurden jedoch keine eigenständigen Methoden mehr entwickelt. Wir werden im nächsten Abschnitt noch einmal auf eine Arbeit von M. Friedman zurückkommen.

²⁴⁸ Bezeichnend für den Siegeszug ist die Haltung von G. Tintner. Nach Tintner (1949) bedurfte die Ökonomie einer Möglichkeit, die Wahrscheinlichkeit einer Theorie zu beurteilen: "Recent work in econometrics and economic statistics, especially some of the work done by members of the Cowles Commission, makes it likely that in the not too distant future there may be a fresh interest in questions arising relating to the *probability of a theory*. On the basis of certain empirical evidence we may, for instance, ask about the probability of the Keynesian as compared to the 'classical' economic system." (S. 251). Hervorhebung im Original. Er stellte fest, daß die bisherigen Wahrscheinlichkeitsauffassungen: Häufigkeitsinterpretationen, wie z. B. von Haavelmo (1944) ("mostly rather vaguely"), logische Wahrscheinlichkeiten und Inferenzkonzeptionen, Fishers Fiduzialtheorie, die Neyman-Pearson-Theorie oder Walds Entscheidungstheorie nicht zufriedenstellend seien: "What is desired is a 'pure' theory of induction which would permit us to evaluate ultimately the degree of confirmation of a given statistical hypothesis without any pragmatic considerations, as implied in Wald's theory. This is particularly important in economic statistics, and more generally in the statistical practice relating to social phenomena." (S. 252). Diese Theorie der Induktion sah er von R. Carnap verwirklicht. Elf Jahre später konstatierte er hinsichtlich einer solchen Beurteilung (volkswirtschaftlicher) Hypothesen anhand von Bestätigungsgraden: "Das ist im Prinzip mit Hilfe der Carnapschen Theorie möglich, aber man muß gestehen, daß die Theorie noch nicht so weit ausgearbeitet ist, um sich mit so komplizierten Problemen zu befassen." Tintner (1960, S. 17). Statt dessen zog er nun die Häufigkeitsinterpretation bzw. Stichprobenkonzeption vor: "Die relativen Häufigkeiten können auch aufgefaßt werden als Stichproben aus einer unendlichen Grundgesamtheit, die alle möglichen und vorstellbaren Volkswirtschaften umfaßt usw." (S. 17). Auf S. 40ff folgte dann explizit eine Empfehlung der Theorie von Neyman/Pearson.

Aufgabe zukünftiger ökonometrischer Forschung in einem "social engineering".²⁴⁹

"The basic principles of the statistical analysis of *systems* of relationships (such as supply and demand equations) have been revised [...]. The traditional method of least squares [...] must be replaced by certain other methods when the problem is one of 'social engineering' (advice to firms, government agencies)."²⁵⁰

Die Grundlagen des Haavelmo-Ansatzes wurden von der *Cowles Commission* nicht mehr hinterfragt. Ein 1950 unter dem Titel "*Statistical inference in economics*" veröffentlichter Beitrag von J. Marschak behandelte bereits nur noch (schätz) technische Probleme innerhalb des vorgegebenen Rahmens.²⁵¹

Welche Bedeutung kommt nun dem Aspekt der Modellbildung in der sich entwickelnden "Mainstream"-Ökonometrie zu? Nach Qin war das Interesse in den ersten Jahren der Ökonometrie völlig auf die Formalisierung der Schritte 'Schätzung', 'Identifikation' und 'Testen' ausgerichtet, die Rolle von "Modellen" in diesem Zusammenhang dagegen nicht behandelt worden.²⁵² Aufgrund der offensichtlichen Diskrepanzen zwischen angestrebten Zielen und anhand konkreter Daten erzielter Ergebnisse habe man dann aber zwei Problemkreise erkannt: zum einen, daß zwischen statistischen und ökonomischen Gesetzen unterschieden werden müsse, zum anderen, daß dem Aspekt der statistischen Inferenz stärkere Beachtung geschenkt werden sollte. Weiterhin habe man gesehen, daß es fruchtbarer sei, "Modelle" statt "Gesetze" als Zielgrößen für empirische Untersuchungen zu propagieren: "Models thus made up the substance of econometrics."²⁵³

Dabei sei nach Qin schon früh zwischen theoretischen Modellen, die aus a priori gegebenen Theorien abgeleitet wurden und statistischen Modellen, die als Ergebnis eines "data-fitting" entstanden waren, unterschieden worden. Und: "It was soon recognized that most theoretical models were incomplete for econometric purposes, whereas statistical models could turn out nonsensical with respect to economic interpretation."²⁵⁴

²⁴⁹ Jacob Marschak [1898-1977] wurde in Kiew (Rußland) geboren, studierte dort von 1915 bis 1918 Maschinenbau, anschließend in Berlin und Heidelberg Ökonomie und Statistik bei L. von Bortkiewicz, bei dem er auch 1922 promovierte, hatte danach mehrere Positionen an Forschungsinstituten und Universitäten in Deutschland inne und schrieb Artikel für die *Frankfurter Allgemeine Zeitung*. 1933 ging er nach England, wurde 1935 Direktor des *Oxford Institute of Statistics* und siedelte 1939 in die USA über. 1946 folgte er J. M. Keynes als Präsident der *Econometric Society*.

²⁵⁰ In einem Report an die *Rockefeller Foundation* 1943, zitiert nach Epstein (1987, S. 61).

²⁵¹ Marschak (1950). Das wichtigste technische Problem, die Parameterverteilung in einem System von linearen stochastischen Differenzengleichungen, wurde bereits 1943 von Mann und Wald gelöst. Vgl. Mann/Wald (1943). Die ersten Lehrbücher der Ökonometrie basierten auf dem Haavelmo-Ansatz, ebenso wie die Einführung von Menges (1959b).

²⁵² Vgl. Qin (1993, S. 37ff).

²⁵³ Ebda., S. 37.

²⁵⁴ Wenn sie anschließend jedoch behauptet: "Consequently there emerged a structural modeling procedure, in which model building was schematized into a process of starting from a structural given by theory and then estimating the structural coefficients from data", so ist dies aber wohl eher ein Widerspruch als eine Konsequenz.

Ihrer Meinung nach habe das Konzept der "Modelle" bereits in den frühen dreißiger Jahren Einzug in die Ökonometrie gefunden. Vor diesem Zeitpunkt habe das Interesse vor allem der statistischen Untersuchung von Nachfrage- und Produktionsfunktionen gegolten. Speziell der Versuch einer statistischen Überprüfung von Postulaten der neoklassischen Theorie habe zwei Schwächen dieser Theorie verdeutlicht: die starke Abhängigkeit von unrealistischen A-priori-Annahmen und die generelle Vernachlässigung einer dynamischen Perspektive. Konsequenterweise habe sich die Konzeption ökonometrischer Studien daher in diesen Punkten von der vorherrschenden ökonomischen Theorie gelöst.

Hinsichtlich empirischer Untersuchungen auf dem Gebiet der Nachfrage-Analyse stellten sich ebenfalls zwei Probleme: die Auswahl der Variablen und die Wahl der funktionalen Beziehungen, wobei anfangs ökonomische Theorien zugrundegelegt wurden (allerdings um den Faktor "Zeit" erweitert) und man mittels Anpassungsverfahren schrittweise die Anzahl der Variablen erhöhte.²⁵⁵ Solch eine Variablen-selektionsprozedur wurde auch von Moore und Schultz angewendet und durch Frischs "bunch-maps" formalisiert.²⁵⁶ Qin weist zu Recht darauf hin, daß man in diesem frühen Stadium bereits zu einer sehr pragmatischen Haltung überging. Man wählte diejenige Form des funktionalen Zusammenhanges, die aufgrund einer Inspektion der gegebenen Daten als die plausibelste schien. Damit stellte sich natürlich unmittelbar das Problem der Rechtfertigung, falls diese spezifischen Zusammenhänge nicht mehr in Einklang mit ökonomischen Theorien standen. Ebenso wichtig war das Problem des Zeithorizonts der zu untersuchenden Beziehungen. War das Ziel die Ermittlung von Koeffizienten, die über langfristige oder kurzfristige Beziehungen Auskunft geben sollten?

An dieser Vorgehensweise änderte sich *prinzipiell* wenig, als die *Cowles Commission* das Programm Haavelmos umsetzte.²⁵⁷ Bemerkenswert ist, daß diese Entwicklung zum großen Teil Keynes zu verdanken ist. Seine 1936 veröffentlichte *General Theory* bildete die Grundlage für eine ganze Generation ökonometrischer Modelle.²⁵⁸ Mit den Jahren wuchs auch die Anzahl der berücksichtigten Variablen und Gleichungen. Die ersten Modelle sind vor allem mit dem Namen Lawrence Klein verbunden, der die von Haavelmo eingeführten indirekten Kleinstquadratschätzungen verwendete.²⁵⁹ Zunächst benutzte er ein von Haavelmo 1947 konstruiertes Modell eines Nationaleinkommenssystems mit zwei Gleichungen, in dem er die Konsumfunktion um das eine Periode zurückliegende Realeinkommen

²⁵⁵ Qin (1993, S. 40): "This implied a simple → general approach."

²⁵⁶ Vgl. ebda., S. 47.

²⁵⁷ Vgl. zum folgenden Wallis (1994b). Einen ausführlichen Überblick über die historische Entwicklung der makroökonometrischen Modellbildung geben Bodkin/Klein/Marvah (1991).

²⁵⁸ Kirchgässner (1983, S. 513): "Es ist eine Ironie der Wissenschaftsgeschichte, daß ausgerechnet die keynesianische Theorie der Ökonometrie zu diesem Aufschwung verhalf, obwohl J. M. Keynes selbst [...] davon überhaupt nichts gehalten hat." Dennoch hat Keynes auch institutionell die Entwicklung der Ökonometrie beeinflusst, indem er die Einrichtung eines Departments für angewandte Ökonomie in Cambridge initiierte, dessen erster Direktor R. Stone wurde. Später arbeiteten dort J. Durbin und G. S. Watson. Vgl. dazu King (1992).

²⁵⁹ Für die entsprechende Literatur sei auf Wallis (1994a) und Christ (1994a) verwiesen.

pro Kopf erweiterte (6 Gleichungen: 3 zu schätzende, 3 Identitäten). Modell II erweiterte dieses Modell zusätzlich um die zurückliegenden *real per capita liquid assets*. Das daran anknüpfende Modell III enthielt dann 12 Gleichungen und vier Identitäten und wurde mit der 'method of reduced forms' geschätzt (heute unter der Bezeichnung *limited information maximum likelihood* geläufig). Alle Modelle umfaßten die Periode von 1921 bis 1941.²⁶⁰

Die zweite Generation makroökonomischer Modelle begann 1955 mit dem Klein-Goldberger-Modell, das die Periode von 1929 bis 1952 mit Ausnahme der letzten drei Kriegsjahre umfaßte.²⁶¹ Als M. Nerlove 1966 einen Überblick über die vorhandenen Modelle gab,²⁶² konnte er immerhin schon 25 verschiedene Modelle für 9 Länder ausfindig machen.

Anschließend wurden die Modelle nochmals erweitert (etwa um Finanz-, Regierungs- und Außenhandelssektoren).²⁶³ Eine der bekanntesten Entwicklungen ist das von J. S. Duesenberry, L. R. Klein und anderen vorangetriebene *Brookings Model Project*, das jedoch keine breite praktische Anwendung fand, sondern statt dessen die Grundlage für methodologische Weiterentwicklungen bildete.²⁶⁴ Ein Ziel dieser Modelle war die sogenannte stochastische Simulation, die die bedingte Verteilung der endogenen Variablen liefern sollte, um die Auswirkungen von Veränderungen der exogenen Variablen zu evaluieren.²⁶⁵ Ein Problem ergab sich bei mehreren Modellen dieser Zeit dadurch, daß die Anzahl der Gleichungen die Anzahl der Fälle überstieg.

Die Konstruktion makroökonomischer Modelle bildete zweifellos die Hauptentwicklungslinie der Ökonometrie. Statistische Grundsatzüberlegungen traten dabei in den Hintergrund. Auf der anderen Seite hat es hierzu aber auch eine alternative Entwicklung gegeben, die mit dem Namen H. Wold verbunden ist und hier nicht unerwähnt bleiben soll.

Wold's Ansatz der Analyse univariater stochastischer Prozesse von 1938 war bereits bei Frisch auf Ablehnung gestoßen.²⁶⁶ Anders als Koopmans 1937 war Wold stärker am statistischen Aspekt als an der Analyse realer ökonomischer Daten

²⁶⁰ Die Analyse von Zyklen spielte in diesem Zusammenhang keine Rolle mehr.

²⁶¹ Die *Cowles Commission* hatte sich zu diesem Zeitpunkt bereits anderen (nicht-ökonomischen) Problembereichen zugewandt.

²⁶² Nerlove (1966).

²⁶³ Eine interessante Nebenentwicklung bilden Überprüfungen/Untersuchungen der *fix-price models with quantity rationing*, in denen unterschiedliche Regime mit klassischer oder keynesianischer Arbeitslosigkeit, unterdrückter Inflation oder Unterkonsumption entstehen können. So untersuchten etwa Artus/Laroque/Michel (1984) dieses Modell für die französische Wirtschaft von 1963 bis 1978: "The historical probabilities that the economy was in the various regimes are calculated, and a standard policy simulation is successfully adapted to the new kind of model." Wallis (1994b, S. xvii).

²⁶⁴ Siehe dazu McCarthy (1992). In diesem Zusammenhang wurden auch bessere Schätzverfahren benötigt. Eine erhebliche rechentechnische Verbesserung wurde Ende der sechziger Jahre durch die Entwicklung des sogenannten Gauss-Seidel-Algorithmus erreicht, der heute noch in zahlreichen Varianten breite Verwendung findet.

²⁶⁵ In diesem Rahmen wurde 1969 von Klein auch eine *counterfactual history*-Untersuchung der Steuerreform von 1964 durchgeführt. Vgl. dazu Wallis (1994b, S. xxi).

²⁶⁶ Siehe oben, Anm. 167.

interessiert. Haavelmo, der in seinem "probability-approach" immer wieder darauf hinwies, daß ökonomische Daten in der Regel Zeitreihen darstellten und auch für diese Wahrscheinlichkeitsgesetze gelten konnten, rezipierte Wolds Arbeit ebenso wenig wie später die *Cowles Commission*.

In einer Reaktion auf Haavelmos Propagierung *simultaner* Gleichungssysteme forderte Wold, statt dessen "rekursive Ketten" zu betrachten. Damit waren Gleichungen gemeint, die nur zeitverzögerte Regressoren enthielten.²⁶⁷ In einem solchen Ansatz konnte das Gleichungssystem auf ein System stochastischer, linearer Differenzgleichungen zurückgeführt werden, das nur noch aus autoregressiven Gleichungen bestand. Ein solches Gleichungssystem konnte mit der herkömmlichen Kleinstquadratmethode "unverzerrt" geschätzt werden. 1956 ging Wold einen Schritt weiter und propagierte seinen Ansatz nun auch aus prinzipiellen Erwägungen.²⁶⁸ Zum einen waren seiner Ansicht nach in simultanen Systemen keine Kausalaussagen möglich, die Modelle könnten somit allenfalls Beschreibungen liefern,²⁶⁹ zum anderen sprach er sich gegen die Verwendung des Fisherschen Maximum-Likelihood-Ansatzes sowie die Neyman-Pearson-Testtheorie aus, da ökonomische Daten keine "Stichproben" darstellten, sondern nicht-experimentelle und passive Größen. Statt dessen sollte das Modell mit Daten anderer Regionen oder Zeiten getestet, dem Vergleich mit den a priori angeführten Überlegungen ausgesetzt, und vor allem sollte die Prognosefähigkeit eines Modells als Kriterium herangezogen werden.

Wolds Ansatz konnte sich jedoch nicht durchsetzen. Die Ökonometrie wurde von den makroökonomischen keynesianischen (statischen) Modellen von Klein und anderen dominiert. Erst Jahrzehnte später wurde Wolds Ansatz in einer grundlegenden Kritik von C. Sims wieder aufgegriffen.²⁷⁰

Methodisch waren die Grundlagen des "mainstream" durch die Arbeiten der *Cowles Commission* gelegt worden. In diesem Rahmen blieben auch die Anwendungen von Klein und anderen, auch wenn man aufgrund der immensen technischen Probleme des vollständigen Maximum-Likelihood-Ansatzes wieder auf Kleinstquadratschätzungen zurückgriff. Zwar flossen über zeitreihenanalytische Kritiken immer wieder technische Verfeinerungen ein,²⁷¹ doch reine Zeitreihenansätze fanden keine große Beachtung mehr.²⁷² Erst durch die Arbeit von Box/Jen-

²⁶⁷ Vgl. Bentzel/Wold (1946). Die (wirtschafts-) theoretische Fundierung lieferte Wold die sog. "Stockholmer Schule". Vgl. zum folgenden auch Morgan (1991).

²⁶⁸ Vgl. Wold (1956). Verschiedene Vergleiche hatten überdies gezeigt, daß in realen Situationen eine Kleinstquadratschätzung nicht schlechter abschnitt (bezogen auf theoretische Größen wie Standardfehler oder "bias" sowie hinsichtlich ihrer Prognosefähigkeit) als eine Maximum-Likelihood-Schätzung.

²⁶⁹ "There were major differences between Haavelmo and myself. He started with interdependent systems and used maximum likelihood – which was not good. I used recursive systems and least squares. But in any case, he confused interdependent systems and simultaneous systems." Wold in Hendry/Morgan (1994, S. 423).

²⁷⁰ Siehe dazu den nächsten Abschnitt.

²⁷¹ Z. B. durch Cochrane/Orcutt (1949) oder Durbin/Watson (1950), Durbin/Watson (1951).

²⁷² Wir hatten in Abschnitt A 4 b darauf hingewiesen, daß die Entwicklung der Theorie stochastischer Prozesse seit den fünfziger Jahren nur noch rein mathematischer Natur war

kins²⁷³ änderte sich das Bild. Zu Beginn der siebziger Jahre geriet die makroökonometrische Modellbildung in eine Krise. Zu dieser Zeit stellte man fest, daß Prognosen "naiver" ARIMA-Modelle bessere Resultate lieferten als die umfangreichen makroökonometrischen Modelle.²⁷⁴ Mit der zunehmenden Akzeptanz der Box-Jenkins-Methodologie verschwommen dann die Grenzen zwischen Zeitreihenanalyse und Ökonometrie.²⁷⁵

e. Ökonometrische Methodologien

Im letzten Abschnitt klang ein Problem an, das bereits bei der Diskussion der Konzepte der Statistik als zentral angesehen wurde: die Abhängigkeit inferentieller Aussagen von der Wahl des Modells, speziell von der Wahl des Modells aufgrund eines "passiven" Datensatzes. Nicht nur im Rahmen der umfangreichen makroökonometrischen Modelle seit Tinbergen stellte sich diese Problematik, sondern auch schon im einfachsten Fall einer bivariaten Regression. Ein Zitat von F. M. Fisher belegt anschaulich das Problem:²⁷⁶

"Consider the case of the econometrician who plots a scatter diagram or performs a regression on a set of observations and then notices that a few of the points seem to be off the main scatter or off the regression line in such a marked and similar manner as to make him suspect that in the years to which those points correspond there were, in fact, parameter shifts or exogenous shocks – that mechanisms other than that which he is investigating generated those observations. He then looks deeper into the literature or into the history of the period and finds convincing arguments that such was indeed the case. If he retains the observations, he is deliberately

und keine Verbindung mehr zu ökonomischen Problemen gesehen wurden, wie dies noch bei Wold zumindest in Beispielen der Fall war. Für die *Cowles Commission*, damals noch im Rahmen eines Projektes zur Aktienkursanalyse, hatte Davis (1941) eine umfangreiche Monographie über die univariate Analyse von Zeitreihen verfaßt, die jedoch nach Haavelmos Stukturgleichungsansatz (der wie erwähnt in vervielfältigter Form bereits 1941 erschienen war) keine Rezeption fand. Eine ebenfalls univariate (autoregressive) Analyse von Preisreihen von M. Kendall (einem Schüler von G. U. Yule) stieß unter den Ökonomen in der Diskussion auf Ablehnung. So klagte etwa R. G. D. Allen, daß seit Yule Ökonomen Zeitreihenkorrelationen nur mit großer Skepsis behandelten, nun aber durch die univariate Betrachtungsweise das eigentlich Interessante verloren gehe: "They have been told to beware of a fearsome devil – serial correlation – and to look to statisticians to cast the devil out for them. It seems now that the more statisticians work on the casting out of devils, the more they cast out everything else, including what the economist want." Allen in Kendall (1953, S. 25f). Noch stärker waren die Einwände von H. S. Houthakker, der Kendalls univariate Vorgehensweise mit der längst überwundenen traditionellen Zeitreihenanalyse gleichsetzte: "From about 1930 the development of econometrics has led to a general conviction that little purpose is served by studying individual time series separately, and that they become interesting only when linked to each other in economically meaningful relations." Ebda., S. 32. Die *Spektralanalyse*, die im Gegensatz zur Periodogrammanalyse der zwanziger Jahre explizit die Eigenschaften stochastischer Prozesse im Frequenzbereich untersuchte, erlebte dagegen in den 60er Jahren eine – wenn auch kurze – Blüte. Vgl. dazu etwa Granger (1964). Unter den 50 von Harvey (1994a) zusammengestellten zeitreihenanalytischen Arbeiten findet sich dagegen keine einzige spektralanalytische Untersuchung.

²⁷³ Siehe oben, Abschnitt A 4 b.

²⁷⁴ Siehe vor allem Nelson (1972).

²⁷⁵ Unterschiede und Gemeinsamkeiten bei Zellner/Palm (1974) oder Wolters (1987).

²⁷⁶ Fisher (1962, S. 7). Hervorhebung von uns.

retaining a bias of unknown extent in his results; if he discards them in accordance with the principle of selective estimation, it will at the very least be unclear what meaning should be attached to his confidence intervals, since it will be difficult to interpret precisely the statement that if he had gone through the same procedure one hundred times, the derived interval would be expected not to cover the true parameter five of those estimates. *The point is that it is not at all clear what is meant by 'the same procedure' in these circumstances.*"

Nicht nur eine solche informelle, visuelle Inspektion der Daten stellt inferenzstatistische Aussagen in Frage. Um die oftmals enttäuschende Anpassung der Modelle zu verbessern, fanden in der Ökonometrie zunehmend Variablenselektionsverfahren Verbreitung. In welchem *Ausmaß* dabei Verzerrungen entstehen können, wollen wir hier anhand der bereits in Abschnitt A 4 d angesprochenen Arbeit von M. C. Lovell illustrieren. Lovell hat einen interessanten heuristischen Lösungsweg eingeschlagen. Er ging von einer *fiktiven* Zielgröße Y aus, die von *realen* potentiellen Regressoren X_i und einer *fiktiven* Störgröße erzeugt wurde. Dadurch konnte man die Auswirkungen von Selektionsmechanismen anhand realer Daten untersuchen und hatte dennoch die Möglichkeit wiederholter Versuche. In seiner Erhebung ging Lovell von einer fiktiven Konsumreihe aus.

Tab. 4: Mögliche erklärende Variablen für die fiktive Entwicklung des Konsums, 1948-1970.

Nr.	Typ	Name
1		Index, bestehend aus 5 Indikatoren
2		Implicit price deflator des Bruttosozialprodukts
3	F	Ausgaben der Regierung (Güter und Dienstleistungen)
4	F	Ausgaben der Bundesregierung (Güter und Dienstleistungen)
5	F	Anschaffungen der Bundesregierung, NPA
6		Bruttosozialprodukt
7		Potential Level of GNP in Preisen von 1958 (\$)
8		Verfügbares Einkommen
9		Erwartete Investitionsausgaben
10	M	Bankreserven
11	M	Monetary Base (St. Louis FED Concept)
12	M	Geldmenge M1
13	M	Geldmenge M2
14		Dow-Jones industrielle Aktienpreise
15		AAA Corporate Bond Yield (Moody's)
16		Arbeitsangebot
17		Arbeitslosenrate
18		Unfilled Orders, MFG Durable Goods Industries
19		Neue Aufträge, Manufacturing industries
20		Trendvariable

F: fiskale Variablen, M: monetäre Variablen, N = 23

Als potentielle Kandidaten, die in einem Selektionsverfahren als mögliche erklärende Variablen des Konsums ausgewählt werden könnten, wählte er 20 Variablen aus den über 3000 der Datenbank des National Bureau of Economic Research aus. Für die (fiktive) Entwicklung des Konsums könnten nach Ansicht Lovells die in Tab. 4 wiedergegebenen 20 Variablen in Frage kommen. Diese Variablen werden als potentielle Kandidaten angesehen, um neun artifizielle abhängige Variablen zu "erklären".²⁷⁷

Tab. 5: Konstruierte Modelle, die die artifizielle Konsumreihe generieren. Quelle: Lovell (1983, S. 6).

Nr.	Modell	
I	$Y_{1,t} = 130 \epsilon_t$	
II	$Y_{2,t} = 130 \epsilon_t^*$	
III	$\log Y_{3,t} = .131 + .865 \log Y_{3,t-1} + .121 \log Y_{3,t-2} + .0016t + .0178 \epsilon_t$ (.898) (.224) (.266) (.0097)	$\frac{R^2}{S_e} = \frac{.997}{.018}$
IV	$Y_{4,t} = -325.0 + 4.44X_{12,t} + 12.55\epsilon_t$ (13.8) (0.09)	$\frac{R^2}{S_e} = \frac{.991}{12.54}$
V	$Y_{5,t} = 75.3 + 2.4 X_{3,t} + 18.58\epsilon_t$ (8.9) (0.7)	$\frac{R^2}{S_e} = \frac{.980}{18.58}$
VI	$Y_{6,t} = -125.0 + 2.22X_{12,t} + 1.2 X_{3,t} + 15.56\epsilon_t$	
VII	$Y_{7,t} = -325.0 + 4.44X_{12,t} + 12.55\epsilon_t^*$	
VIII	$Y_{8,t} = 75.3 + 2.4 X_{3,t} + 18.58\epsilon_t^*$	
IX	$Y_{9,t} = -125 + 2.22X_{12,t} + 1.2X_{3,t} + 15.56\epsilon_t^*$	
$\epsilon_t \sim N(0, 1) \quad \epsilon_t^* = 0.75 \epsilon_{t-1}^* + \epsilon_t \sqrt{7/4}$		

Die Parameter der Modelle III, IV und V wurden durch eine Regression anhand der tatsächlichen Konsumausgaben pro Kopf ermittelt. Die Parameter von Modell VI sind Durchschnitte der Modelle IV und V. Modell VII, VIII und IX haben dieselben Parameter wie Modell IV, V und VI, jedoch zusätzlich einen autokorrelierten Fehlerterm.

Für jedes dieser Modelle wurden 50 Realisationen erzeugt und anschließend anhand von drei Verfahren aus den 20 Variablen aus Tab. 4 jeweils die beiden Variablen ausgesucht, die für die entsprechenden Daten $Y_{i,j,t}$ ($i = 1, \dots, 9; j = 1, \dots, 50; t = 1948, \dots, 1970$) die besten Anpassungen lieferten. Diese drei Kriterien sind

1. die Variablen, die in einer schrittweisen Regression ausgewählt werden,
2. das maximale \bar{R}^2 und
3. $\max - \min |t|$.

Wenn man zunächst ignoriert, welche Variablen ausgewählt wurden und nur den "Erklärungsanteil" der jeweils zwei besten Regressoren betrachtet, so stellt man fest, daß abgesehen von Modell I und II die Varianz von Y durch die ausgewählten Variablen nahezu vollständig erklärt wird: \bar{R}^2 liegt zwischen 0.982 (Modell V) und 0.998 (Modell III).²⁷⁸ Es sollten also gute Chancen bestehen, daß das richtige

²⁷⁷ Auf die zahlreichen weiteren – auch von Lovell nicht behandelten – schätztheoretischen Probleme, die z. B. mit der Nichtstationarität verbunden sind, gehen wir nicht ein, da wir hier nur einen Aspekt illustrieren wollen.

²⁷⁸ Siehe ebda., S. 7.

Modell gefunden wird. Sind nun die *data-mining*-Verfahren in der Lage, die richtigen Variablen auszuwählen? Betrachten wir, welche Variablen für die einzelnen Modelle jeweils ausgewählt wurden: Tab. 6 zeigt, wie häufig jede Variable anhand der einzelnen Verfahren in den 50 Simulationen ausgewählt wurde, wenn die Daten aufgrund von Modell II, IV oder IX simuliert wurden.²⁷⁹

Tab. 6: Durch "data mining" ausgewählte Variablen (Auswahl) - Häufigkeiten einzelner Variablen (50 Simulationen).

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Modell III																				
Schrittweise	0	3	1	0	2	12	16	8	0	2	0	4	12	1	12	18	0	5	0	4
Maximales \bar{R}^2	0	2	0	0	0	10	24	2	0	0	0	0	18	5	14	5	1	8	0	11
max - min $ t $	1	2	5	5	0	1	0	0	0	3	1	0	35	7	4	2	1	1	0	32
Modell VI																				
Schrittweise	3	1	28	12	7	3	1	1	2	2	5	28	1	1	1	0	0	3	0	1
Maximales \bar{R}^2	4	2	19	17	4	1	6	1	1	2	6	26	1	1	2	1	0	5	0	1
max - min $ t $	0	0	0	1	0	0	0	0	0	48	0	0	0	2	2	0	0	0	1	46
Modell IX																				
Schrittweise	1	6	26	7	2	2	0	2	3	0	13	25	1	0	2	3	0	4	1	2
Maximales \bar{R}^2	1	3	12	17	1	4	5	1	1	0	14	23	1	0	5	6	0	5	0	1
max - min $ t $	0	0	1	6	0	0	1	1	0	43	0	0	0	3	4	0	0	1	0	40

Liegt Modell III zugrunde, ist die einzige "exogene" Variable die Trendvariable (20). Die schrittweise Regression erkannte dies in 50 Simulationen lediglich viermal, in 18 Fällen wurde irrtümlich Variable 16 (Arbeitsangebot), in 16 Fällen Variable 7 (GNP) ausgewählt. Anhand des zweiten Kriteriums (Maximales \bar{R}^2) wurde in 11 Fällen die Trendvariable selektiert, in 24 Fällen irrtümlich Variable 7. Bei Modell VI wurde in etwa der Hälfte der Fälle durch die schrittweise Regression jeweils eine richtige Variable (3 oder 12) erkannt, durch das zweite Kriterium Variable 3 in 19 Fällen, Variable 12 in 26 Fällen; mit Hilfe des dritten Kriteriums wurden die richtigen Variablen kein einziges Mal erkannt. Ähnliche Größenordnungen ergeben sich für Modell IX. Nach diesem Kriterium wurde in mindestens 80 % der Simulationen irrtümlich Variable 10 (Bankreserven) oder die Trendvariable ausgewählt.

Zusammenfassend gelangte Lovell zu dem Schluß, daß anhand der schrittweisen Regression in etwa 70 % der Fälle die korrekte Variable, in etwa 82 % die korrekte oder eine "ähnliche" ausgewählt wird, die Wahrscheinlichkeit eines Fehlers erster Art (das "wahre" Signifikanzniveau) 30 % beträgt, der Fehler zweiter Art 15 %. Die anderen beiden Selektionsverfahren schneiden noch schlechter ab: Durch die Maximierung des Bestimmtheitsmaßes werden lediglich in 52 % der Fälle die richtigen Variablen ausgewählt, anhand des max-min $|t|$ Verfahrens in keinem

²⁷⁹ Quelle: ebda., S. 9. Dort findet sich auch eine vollständige Tabelle. Strenggenommen müßte man die Anzahl der Kombinationen der ausgewählten Variablen betrachten, doch dann wäre das Problem kaum noch überschaubar.

einzigsten Fall. Die Wahrscheinlichkeit eines Fehlers erster Art ist erheblich höher: 53 % und sogar 81 %. Lediglich die Fehler zweiter Art (die Wahrscheinlichkeit, die Nullhypothese nicht abzulehnen, wenn sie falsch ist) sind mit 8 % bzw. 0 % geringer.²⁸⁰ Lovell kommentierte dies wie folgt: "It is ironic that the data mining procedure that is most likely to produce regression results that appear impressive in terms of the customary criteria [das max-min $|t|$ Kriterium, T. R.] is also likely to be the most misleading in terms of what it asserts about the underlying process generating the data under study."²⁸¹ Demnach könnten t -Werte bei einem solchen Vorgehen allenfalls deskriptiven Charakter haben; bei Tests sollte daher erheblich konservativer vorgegangen werden.

Lovell hat mit dieser Kritik keine Forderung nach einem grundsätzlichen Umdenken verbunden oder gar eine Methodologie begründen wollen, sondern lediglich auf ein Problem aufmerksam gemacht, das sich generell bei der Analyse einer nicht variierbaren Anzahl von Daten stellt. Letztendlich konnte er – neben dem Aufruf an alle empirisch arbeitenden Wissenschaftler, den Modellselektionsprozeß vollständig mitzuteilen – auch keine andere Lösung anbieten, als sie schon Tukey und Friedman forderten: die *crossvalidation*, also den Datensatz in einen Modellgenerierungs- und einen Validierungsteil aufzuteilen, eine Forderung, die bei der schmalen Datengrundlage in der Regel illusorisch sein dürfte.

Dieses Grundproblem stellt sich der Ökonometrie als der "Pionierwissenschaft auf dem Gebiet der nichtexperimentellen Modellbildung."²⁸² Die Ökonometrie ist im weiteren Verlauf aber nicht – wie man vielleicht aufgrund einer Lektüre der meisten Lehrbücher vermuten könnte – einem einheitlichen Weg gefolgt, sondern hat sich in mehrere Methodologien "aufgelöst", die gegenwärtig miteinander konkurrieren.²⁸³

Zuerst ist in diesem Zusammenhang die Methodologie C. Sims zu nennen, der den mit zahlreichen (insbesondere Identifizierbarkeits-) Annahmen verbundenen Modellen der klassischen Ökonometrie vektorautoregressive Modelle entgegenstellte, deren Struktur und Parameter anhand der Daten ermittelt werden sollten. Diese Modelle haben ihre Tradition in ökonometrischer Hinsicht in den Arbeiten H. O. A. Wolds (*causal-chain-approach*), in statistischer Hinsicht in einer Verallgemeinerung der univariaten Box-Jenkins-Methodologie.²⁸⁴

²⁸⁰ Vgl. ebda., S. 10f.

²⁸¹ Ebda., S. 10. Er betonte, daß hier zwar lediglich 23 Beobachtungen zugrundegelegt worden seien, dafür habe er aber auch einen sehr "günstigen" Fehlerterm unterstellt. Eine Simulation mit einer verdoppelten Varianz des Fehlerterms habe zu dramatisch schlechteren Resultaten geführt. Man sollte auch bedenken, daß hier die funktionalen Formen bereits vorgegeben sind.

²⁸² So der Titel eines Aufsatzes von Wold (1969). Einen Fragenkatalog hinsichtlich des Modellprozesses hat Granger (1990b, S. 14ff) zusammengestellt.

²⁸³ Neben diesen Methodologien hat sich auch eine Radikalkritik entwickelt, die von der "scientific illusion" solcher ökonometrischer Ansätze spricht, wie z. B. Summers (1991) oder auch McCloskey (1990), McCloskey (1993), die wir in diesem Zusammenhang aber nicht näher behandeln wollen.

²⁸⁴ Methodologie insofern, als nicht nur Modelle und Schätztechniken übernommen wurden, sondern auch die Modellanpassungs- und Modellbeurteilungsverfahren.

Auf der anderen Seite hat es auch bayesianische Entwicklungen in der Ökonometrie gegeben, deren herausragendste Vertreter sicher A. Zellner und E. E. Leamer sind. Während Zellner jedoch vorrangig versucht, die klassische Ökonometrie aus bayesianischer Sicht heraus aufzufassen, geht Leamer völlig eigene Wege, die vor allem mit dem Konzept der "Sensitivitätsanalyse" verbunden sind. Nach einem Überblick über die "quantitative" Entwicklung der bayesianischen Ökonometrie wollen wir daher kurz die entscheidenden Elemente der Methodologie Leamers vorstellen.

Die dritte Methodologie ist mit dem Namen D. F. Hendry verbunden und stellt den vorläufigen Höhepunkt einer eigenständigen *englischen* ökonometrischen Entwicklung dar. Hendry bleibt epistemologisch zwar in der Tradition der klassischen (häufigkeitstheoretischen) Inferenz, formuliert aber zahlreiche Ansätze, die dem *data-mining*-Prozeß Rechnung tragen und seiner Ansicht nach zu einer optimalen objektiven Inferenz führen. Diese durch massive Verwendung diagnostischer Tests und eine – noch näher zu charakterisierende – *general-to-specific*-Vorgehensweise gekennzeichnete Methodologie wurde von ihm selbst einmal mit "Prokrustes-Ökonometrie" bezeichnet.²⁸⁵

Alle drei "Methodologien" wurden von ihren Protagonisten in zahlreichen technischen und programmatischen Arbeiten seit Beginn der siebziger Jahre dargelegt, weiterentwickelt und erstmals auf dem fünften ökonometrischen Weltkongreß 1985 gegenübergestellt.²⁸⁶ Bei der Darstellung dieser Methodologien, die ein eigenes Buch füllen könnte, wollen wir uns hier vorerst auf die grundlegenden Prinzipien beschränken. Schließlich wollen wir diesen ökonometrischen Methodologien eine "pragmatische Alternative" gegenüberstellen, die mit dem Namen des berühmten Ökonomen M. Friedman verbunden ist. Friedman hatte bereits 1953 in einer grundlegenden und vielbeachteten Arbeit einen konsequent instrumentalistischen Standpunkt vertreten. Für unseren Zusammenhang ist Friedmans methodologische Position von besonderer Bedeutung, weil er sich zusammen mit A. Schwartz in umfassender Weise historischen Entwicklungen zugewendet hat und schließlich Teilaspekte dieser Untersuchung einer Reanalyse durch Hendry unterzogen worden sind, die somit einen direkten Vergleich zumindest dieser beiden Positionen erlauben.²⁸⁷

Minnesota-Agnostik: C. A. Sims

Die erste hier zu behandelnde Methodologie, die einmal als "Minnesota-Agnostik" bezeichnet wurde, stammt von C. A. Sims.²⁸⁸ Sims folgte dem Weg, der von H. O.

²⁸⁵ Vgl. Hendry/Mizon (1985).

²⁸⁶ Vgl. die Beiträge in Bewley (1987). Diese Gegenüberstellung bildete die Grundlage eines ersten Vergleiches von Pagan (1987), dem ein "update", Pagan (1995), folgte. Weitere Vergleiche versuchen Granger (1990b), Pagan (1987), Pagan (1995) und Hylleberg/Paldam (1991). Eine Reihe von Texten zur methodologischen Grundsatzdiskussion ist in Granger (1990a) wiederabgedruckt. Aufschlußreich ist darüber hinaus eine Diskussion von Hendry/Leamer/Poirier (1990), in der eine Reihe von epistemologischen und methodologischen Fragen im direkten Austausch beantwortet wurde.

²⁸⁷ Siehe Friedman/Schwartz (1982), Hendry/Ericsson (1991) und Friedman/Schwartz (1991).

A. Wold vorgezeichnet wurde. Wie wir im letzten Abschnitt gesehen hatten, lehnte Wold die von Haavelmo propagierte Vorgehensweise ab und schlug statt dessen einen Kausalkettenansatz vor, wobei die Koeffizienten anhand des Kleinstquadratverfahrens geschätzt werden sollten.

Sims griff 1980 eine Kritik von Liu aus dem Jahre 1960 auf. Liu hatte darauf aufmerksam gemacht, daß die Identifikation von simultanen Gleichungssystemen nicht mehr möglich sei, wenn durch Fehlspezifikationen relevante Variablen aus Gleichungen ausgeschlossen wurden. Da seiner Ansicht nach alle ökonometrischen Modelle Approximationen darstellten, könnten in der Regel auch die Restriktionen nur approximativ sein; damit seien die Modelle aber nicht identifizierbar.²⁸⁹ Um dieses Problem zu umgehen, schlägt Sims die Verwendung von sog. "vektorautoregressiven" Modellen vor. Diese Modelle, die technisch eine multivariate Erweiterung der autoregressiven univariaten Prozesse $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p}$ darstellen, sind dadurch gekennzeichnet, daß auf der rechten Seite nur verzögerte Variablen enthalten sind und keine simultanen Beziehungen vorkommen. Für zwei Variablen lautet z. B. ein bivariater AR(2)-Prozeß allgemein:²⁹⁰

$$(3) \quad \begin{aligned} Y_{1,t} &= \phi_{111} Y_{1,t-1} + \phi_{112} Y_{2,t-1} + \phi_{211} Y_{1,t-2} + \phi_{212} Y_{2,t-2} + \varepsilon_{1,t} \\ Y_{2,t} &= \phi_{121} Y_{1,t-1} + \phi_{122} Y_{2,t-1} + \phi_{221} Y_{1,t-2} + \phi_{222} Y_{2,t-2} + \varepsilon_{2,t} \end{aligned}$$

bzw. in Matrixform

$$(4) \quad Y_t = \begin{bmatrix} \phi_{111} & \phi_{112} \\ \phi_{121} & \phi_{122} \end{bmatrix} Y_{t-1} + \begin{bmatrix} \phi_{211} & \phi_{212} \\ \phi_{221} & \phi_{222} \end{bmatrix} Y_{t-2} + \varepsilon_t.$$

Ist $\phi_{112} = \phi_{212} = \phi_{121} = \phi_{221} = 0$, sind beide Prozesse unabhängige, univariate AR(2)-Prozesse, ist $\phi_{112} \neq 0$ und/oder $\phi_{212} \neq 0$, hat Y_2 einen kausalen Einfluß auf Y_1 , ist $\phi_{121} \neq 0$ und/oder $\phi_{221} \neq 0$, hat Y_1 einen kausalen Einfluß auf Y_2 .²⁹¹ Wenn sowohl mindestens ein Koeffizient einer Verzögerung von Y_2 in der ersten Gleichung als auch mindestens ein Koeffizient einer Verzögerung von Y_1 in der zweiten Gleichung ungleich Null ist, entsteht eine "Feedback"-Situation.

Da die "wirkliche" Anzahl der zu berücksichtigenden Verzögerungsterme in der Regel unbekannt sein dürfte, schlug Sims vor, von einem möglichst allgemeinen Modell auszugehen. Zwar können mit solchen Modellen eine Vielzahl von Strukturen modelliert werden, doch ist damit auch eine sehr große Anzahl von Parametern verbunden. So müßten in den von Sims betrachteten 6-Variablen-Modellen²⁹²

²⁸⁸ Der Ausdruck stammt von Johnston (1991, S. 55). Aktuelle Informationen zu Sims' Forschungstätigkeiten sind im Internet auf einem Server der *Yale University* unter der Adresse <http://www.econ.yale.edu/~sims> zu finden.

²⁸⁹ Vgl. Sims (1980a), Liu (1960). Sims bezog sich sowohl auf die Auswahl der Variablen als auch auf die Anzahl der zu berücksichtigenden Verzögerungen.

²⁹⁰ Zu den technischen Einzelheiten solcher Modelle im allgemeinen vgl. ausführlich Lütkepohl (1991).

²⁹¹ Die Definition und Formalisierung einer solchen Situation als "Kausalität" stammt von Granger (1969).

mit 4 verzögerten Termen (z. B. bei Quartalsdaten) 144 Parameter geschätzt werden.²⁹³ Sims forderte daher, die Parameteranzahl zu restringieren. Hierfür befürwortete er auf der einen Seite bayesianische Ansätze, beurteilte aber andererseits die konstruierten Modelle anhand von empirischen Signifikanzniveaus.²⁹⁴ Ein Problem bezüglich dieser vektorautoregressiven Modelle bestehe in der direkten Interpretation der Parameter: "It is especially difficult to make sense of them by examining the coefficients in the regression equations themselves. The estimated coefficients on successive lags tend to oscillate, and there are complicated cross-equation feedbacks. The common econometric practice of summarizing distributed lag relations in terms of their implied long-run equilibrium behavior is quite misleading in these systems."²⁹⁵ Als Alternative führte Sims daher die Betrachtung von "Impulsresponsefunktionen" ein. Diese geben darüber Auskunft, welche Wirkung ein Schock in einer Variable auf den Verlauf dieser Variable selbst und auf den Verlauf der anderen Variablen hat. Wir wollen die Konzeption anhand eines einfachen VAR(2)-Modells illustrieren. In dem Modell

$$(5) \quad Y_t = \begin{bmatrix} 1.14 & 0.4 \\ 0 & 1.34 \end{bmatrix} Y_{t-1} + \begin{bmatrix} -0.5 & 0 \\ 0 & -0.5 \end{bmatrix} Y_{t-2} + \varepsilon_t$$

haben beide Reihen eine zyklische Entwicklung. Zusätzlich übt Y_2 einen Einfluß auf Y_1 über den Koeffizienten $\phi_{112} = 0.4$ aus. Y_2 ist seinerseits dagegen unabhängig von Y_1 ($\phi_{121} = \phi_{221} = 0$).

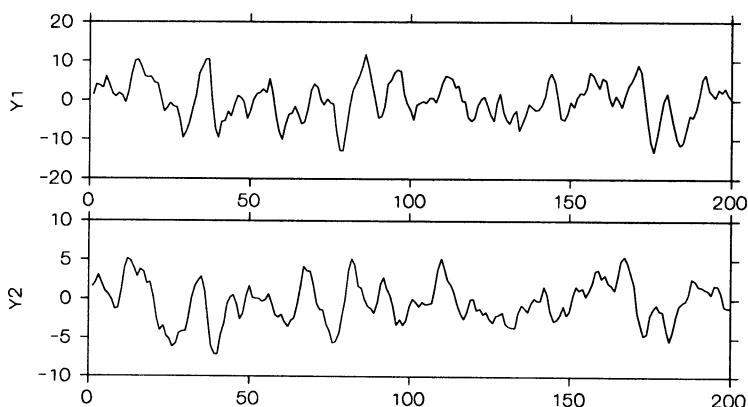


Abb. 11: Realisation eines VAR(2)-Prozesses aus Gleichung (5).

²⁹² Sims untersuchte makroökonomische Modelle für die USA (1949-1975) und Westdeutschland (1958-1976) jeweils mit Variablen zur Messung der Geldmenge, des Bruttosozialprodukts, der Arbeitslosenrate, der Lohn- sowie der allgemeinen und Importpreisentwicklung.

²⁹³ Vgl. Sims (1980a, Abschnitt 2). Wenn jede Variable jede andere bei jeder Verzögerung beeinflussen kann, wächst die Anzahl der Parameter quadratisch mit der Zahl der Variablen.

²⁹⁴ Vgl. Sims (1980a [1990, S. 154, 158]).

²⁹⁵ Ebda., S. 159.

Die aus den in Abb. 11 dargestellten Realisationen geschätzten Impulsantwortfunktionen geben die gegenseitigen Auswirkungen deutlich wieder.²⁹⁶

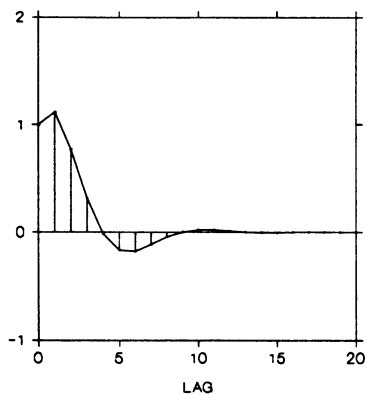


Abb. 12: Reaktion von Y_1 , Schock in Y_1 .

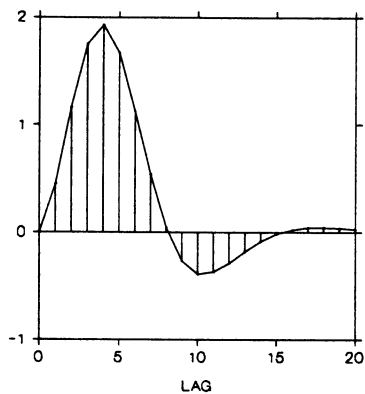


Abb. 13: Reaktion von Y_1 , Schock in Y_2 .

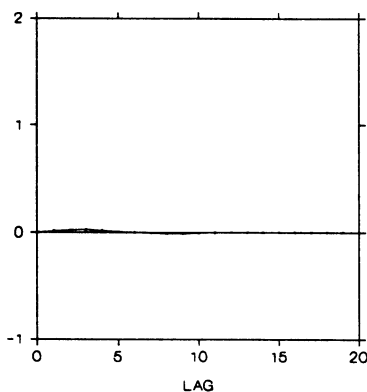


Abb. 14: Reaktion von Y_2 , Schock in Y_1 .

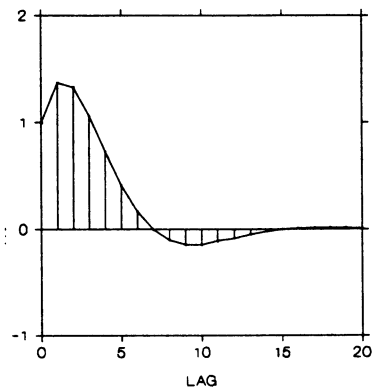


Abb. 15: Reaktion von Y_2 , Schock in Y_2 .

Für 6-Variablen-Systeme, wie sie von Sims spezifiziert wurden, ist mit einer solchen Vorgehensweise die Betrachtung von immerhin jeweils 72 Impulsantwortfunktionen verbunden.²⁹⁷

²⁹⁶ Die Reihen wurden mit zwei unabhängigen Noise-Reihen, $\varepsilon_1 \sim N(0,2)$, $\varepsilon_2 \sim N(0,1)$, simuliert. Ohne den Einfluß von Y_2 hätte die Parameterkombination von Y_1 einen Zyklus von etwa 10 Zeiteinheiten, die Parameterkombination von Y_2 impliziert dagegen einen Zyklus von etwa 20 Zeiteinheiten. Vgl. Rahlf (1996b, S. 25). Die Impulsantwortfunktionen wurden mit dem Programm MULTI berechnet. Siehe dazu Haase et al. (1992).

²⁹⁷ Siehe Sims (1980a [1990, S. 176-187]).

Sims' Vorgehensweise, die schnell Nachfolger fand,²⁹⁸ bedeutete eine Abkehr von der klassischen Ökonometrie.

Zwar mußten die einzubeziehenden Variablen immer noch theoretisch bestimmt werden, die *Struktur* jedoch wurde nicht mehr a priori aufgrund wirtschaftstheoretischer Überlegungen postuliert, sondern ergab sich aus den konkret vorliegenden Daten anhand von statistischen Kriterien, die zum Teil bayesianischer, zum Teil klassischer Herkunft waren.²⁹⁹ Ein solcher "atheoretischer" Ansatz mußte verständlicherweise Kritik hervorrufen. So warfen etwa T. F. Cooley und S. F. LeRoy Sims' Ansatz vor, daß die von ihm angestrebte *kausale* Analyse nicht möglich sei.³⁰⁰ Für eine *Beschreibung* eines (historischen) Datensatzes hielten sie solche Modelle jedoch durchaus für geeignet.³⁰¹

Eine zweite grundlegende Kritik an der Konzeption von Sims, diesmal aus einer bayesianischen Sicht, übte 1985 E. E. Leamer. Bevor wir uns dieser Kritik zuwenden, wollen wir zunächst einen Blick auf die quantitative Größenordnung der bayesianischen ökonometrischen Literatur werfen und anschließend die Eckpfeiler dieser Entwicklung kurz erläutern.

Bayesianische Ökonometrie und der Ansatz von E. E. Leamer

Um einen Überblick über die explizite Anwendung bayesianischer Konzepte und Methoden in der Ökonometrie zu erhalten, verwenden wir als groben, aber dennoch aufschlußreichen Indikator die Literaturdatenbank ECONLIT (Ausgabe 5/95). Die Datenbank umfaßt eine Bibliographie, die Angaben zu 319.960 ökonomischen Publikationen aus über 400 Zeitschriften sowie Sammelbandbeiträgen, Buchrezensionen, Dissertationen und Working Papers enthält, größtenteils in englischer Sprache. Für den Zeitraum von 1987 bis 1989 weisen etwa 25 % der Einträge Abstracts auf, seit 1990 etwa 50 %. Nahezu alle Einträge sind mit Deskriptoren verschlagwortet. Das Klassifikationssystem basiert auf der Systematik des *Journal of Economic Literature* und des *Index of Economic Articles* und enthält über 600

²⁹⁸ Vgl. dazu die in Wallis (1994b, S. xxiii) angegebene Literatur.

²⁹⁹ Diese "Philosophie" hatten wir bereits in Abschnitt A 4 b in der von G. E. P. Box und G. W. Jenkins entwickelten ARIMA-Methodologie besprochen. Sims beruft sich bei seinem Ansatz jedoch, soweit wir sehen, an keiner Stelle auf deren Arbeit.

³⁰⁰ Vgl. Cooley/LeRoy (1985). Sims selbst hatte bei unterschiedlichen Variablenkonstellationen verschiedene Kausalitäten postuliert. So gelangte Sims (1972) zu dem Schluß, daß die Geldmenge das Einkommen kausal beeinflusse und keine anderen Variablen berücksichtigt werden müßten. In Sims (1980a) betrachtete er dagegen zusätzlich Variablen zur Arbeitslosigkeit, zu Löhnen und Importpreisen, in Sims (1982) Zinsen, Inlandspreise, Staatsausgaben und -einnahmen. Die Konsequenzen dieser spezifischen Auswahlen sind keinesfalls gering. In Abhängigkeit von der Auswahl der Variablen ist entweder die Geldmengenänderung ursächlich für Einkommensänderungen (Sims (1972)), oder die Kausalrichtung ist genau umgekehrt, wenn man zusätzlich die Zinsentwicklung berücksichtigt (Sims (1980b)).

³⁰¹ Sims ging später auf diese Kritik ein, führte Identifikationsrestriktionen ein und konstruierte damit "strukturelle" VAR-Modelle. Vgl. dazu Wallis (1994b, S. xxiii). Um den Trendcharakter und langfristige Zusammenhänge ökonomischer Reihen angemessener berücksichtigen zu können, ist der vektorautoregressive Ansatz um das Kointegrationskonzept erweitert worden. Siehe dazu Pagan (1995), als Einführung Thome (1996). Eine Impulsantwortanalyse ist auch für solche Modelle möglich: siehe dazu Lütkepohl/Reimers (1992).

Kategorien. Die Hauptkategorien lauten: *Accounting, Administration, Agriculture, Business Finance and Investment, Consumer Economics, Country Studies, Development, Domestic Monetary and Financial Theory and Institutions, Economic Growth, Fluctuations, General Economics, Health Economics, History, Industrial Organization, Industry Studies, International Economics, Labor, Marketing, Natural Resources, Planning, Population, Quantitative Economic Methods and Data, Systems, Technological Change, Theory, Urban and Regional Economics, Welfare Programs*. Für eine nähere Aufschlüsselung unterteilen wir den erfaßten Zeitraum in die Abschnitte 1969-1979, 1980-1989 und 1990-1995/5 und zählen jeweils die Anzahl der Publikationen, die bayesianisch und "ökonometrisch-nichtbayesianisch" sind.³⁰² Die Anzahl "nichtbayesianischer ökonometrischer" Publikationen wurde anhand einer Volltextsuche "econometri* not bayes*" ermittelt, bayesianische Publikationen durch die Suchangabe "bayes*".³⁰³ Betrachtet man die Ergebnisse, ergibt sich bei aller gebotenen Zurückhaltung dennoch ein recht aufschlußreiches Bild (Tab. 7). So kann etwa (zumindest quantitativ) von einer wachsenden Bedeutung bayesianischer Konzeptionen oder Methoden kaum die Rede sein. Im Zeitraum 1969-1979 stehen 2.958 nichtbayesianischen ökonometrischen Publikationen 194 bayesianische Arbeiten gegenüber. Dieses Verhältnis (etwa 1/15) steigt dann nur leicht an und beträgt 1990-1995/5 etwa 1/11.³⁰⁴

Tab. 7: In ECONLIT 5/95 ergaßte ökonometrische und bayesianische Publikationen.

Zeitraum	Total:	"econometri* not bayes*":			"bayes*":		
	Anzahl	Anzahl	# SWK	# SW	Anzahl	# SWK	# SW
1969-1979	67.056	2.958	221	4.815	194	53	308
1980-1989	150.602	8.072	334	16.091	681	132	1.117
1990-1995/5	102.302	6.595	310	18.668	619	167	1.565
1969-1995/5	319.960	17.625			1.494		

Welche Themen wurden in diesen Arbeiten behandelt? Hierüber gibt – wiederum nur als sehr grober Indikator – eine Auszählung der Schlagwortkategorien (SWK) Aufschluß (Tab. 8).³⁰⁵ Die 2.958 nichtbayesianischen ökonometrischen Publika-

³⁰² Die Verschlagwortung der Einträge wurde 1991 geändert und ein alphanumerisches System eingeführt, das die alte Systematik ablöste. Aus Konsistenzgründen wurde das alte System aber zusätzlich fortgeführt. Um eine künstliche Doppelzählung zu vermeiden, beschränkt sich die Auszählung auch für die Zeit ab 1991 auf die alten Codes.

³⁰³ Wir gehen davon aus, daß diese Arbeiten im weitesten Sinne ökonometrische Aspekte thematisieren, auch wenn der Begriff "bayes*" nicht explizit vorkommen sollte.

³⁰⁴ Der Anteil ökonometrischer Arbeiten an der Gesamtzahl steigt in diesen Zeiträumen von 4.4 % auf etwa 6.4 %.

³⁰⁵ In Klammern davor jeweils der vierstellige Code. Aus Platzgründen wurden hier jeweils nur die 10 häufigsten Nennungen wiedergegeben.

tionen von 1969 bis 1979 wurden von den Bearbeitern 221 Schlagwortkategorien zugeordnet, wobei insgesamt 4.815 Schlagworte vergeben wurden, den 8.072 Publikationen von 1980 bis 1989 wurden 16.091 Schlagworte aus insgesamt 334 Kategorien zugewiesen etc.³⁰⁶

Die Themen im einzelnen: Betrachten wir zunächst die nichtbayesianischen Publikationen. Es überwiegen Nennungen zu Methoden sowie (ökonometrischen) Modellen und dem Aspekt des "Forecasting". Ökonometrie als Überprüfung ökonomischer Theorien ist dagegen von untergeordneter Bedeutung. 1969-1979 untersuchten 134 Arbeiten u. a. das Thema 'General-Equilibrium-and-Disequilibrium-Theory' (Rang 8), 1980-1989 war das Thema auf den 11. Rang, 1990-1995/5 sogar auf den 45. Rang (76 von 18.668 Nennungen bzw. 6.595 Publikationen) abgerutscht. Das Schlagwort 'Macroeconomic-Theory-General' rangierte 1969-1979 auf Rang 19 (43 von 4.815 Nennungen bzw. 2.958 Arbeiten), 1980-1989 gehörte es mit Rang 6 immerhin zu den "top ten", 1990-1995/5 war es dagegen nur noch auf Rang 33 (100 Nennungen) anzutreffen. Zunehmende Bedeutung gewann dagegen eine spezielle Methodenkategorie: 'Time-Series-and-Spectral-Analysis'. 1969-1979 noch auf Rang 12 (72 Nennungen), stieg die Kategorie 1980-1989 auf Rang 5 und war schließlich 1990-1995/5 das wichtigste ökonometrische Thema überhaupt (1813 Arbeiten).³⁰⁷

Bei den bayesianischen Publikationen ist naturgemäß die selbstbekenkende Pauschalkategorie 'Bayesian-Statistics-and-Bayesian-Econometrics' die im gesamten Zeitraum häufigste Nennung.³⁰⁸ 1969-1979 folgte dann, mit weitem Abstand,³⁰⁹ eine Nennung des auch in der klassischen Ökonometrie traditionell bedeutenden Problems 'Distributed-Correlated-Disturbance-Terms; -Inferential-Problems -in-Single-Equation-Models'. Alle weiteren Angaben sind zahlenmäßig so gering, daß wir auf eine weitere Interpretation verzichten wollen. 1980-1989 schien die Situation in gewisser Hinsicht aufgeschlossener. Nun wiesen von 681 bayesianischen Arbeiten immerhin 82 das Schlagwort 'Econometric-and-Statistical-Methods-and-Models-General' auf.³¹⁰ Dies kann man mit einiger Vorsicht als beginnende Liberalisierung bezeichnen.

³⁰⁶ Jeder Eintrag enthält also durchschnittlich 4.815/2.958 oder etwa 1.63 Schlagworte. Der Wert stieg in der Folgezeit leicht an und betrug 1990-1995/5 etwa 2.8 bzw. 2.5 bei den bayesianischen Arbeiten. Dadurch ist die folgende Aufteilung nach Schlagworten bei den späteren Arbeiten etwas differenzierter.

³⁰⁷ Die Kategorien 'Economic-Planning-Policy' und 'Economic-Planning-Theory', die man am ehesten mit den ursprünglichen Zielen der *Econometric Society* verbinden könnte, belegten zu dieser Zeit mit jeweils 17 Nennungen Rang 78. 1969-1979 lag erstere Kategorie immerhin noch auf Rang 24.

³⁰⁸ Ein entsprechendes Schlagwort 'classical statistics' oder 'Neyman-Pearson statistics' existiert nicht.

³⁰⁹ Wir haben hier zwar durchschnittlich 1.59 Schlagworte je Publikation, diese sind aber ungleich verteilt. Von den 194 Publikationen weisen 119 lediglich das Schlagwort 'Bayesian-Statistics-and-Bayesian-Econometrics', die restlichen 75 in der Regel zwei oder drei weitere auf. Offensichtlich erschien in jenen Fällen also eine Kennzeichnung als lediglich "bayesianisch" ausreichend.

³¹⁰ Davon in immerhin 71 Fällen ohne gleichzeitige Nennung der Kategorie 'Bayesian-Statistics-and-Bayesian-Econometrics'.

Tab. 8: In ECONLIT 5/95 erfaßte ökonomische und bayesianische Publikationen:

Ökonometrische nichtbayesianische Publikationen

1969-1979

(2110) Econometric-and-Statistical-Methods-and-Models-General	447
(2120) Construction,-Analysis,-and-Use-of-Econometric-Models	369
(1324) Forecasting-and-Econometric-Models-Theory-and-Methodology	309
(2113) Distributed-Correlated-Disturbance-Terms;-Inferential-Problems-in-Single-Equation-M...	288
(2112) Inferential-Problems-in-Simultaneous-Equation-Systems	235
(1320) Forecasting;-Econometric-Models-General	188
(2114) Econometric-and-Statistical-Methods-and-Models;-Multivariate-Analysis,-Statistical-I...	185
(0210) General-Equilibrium-and-Disequilibrium-Theory	134
(0222) Microeconomic-Theory-Theory-of-the-Household-Consumer-Demand	114
(0250) Social-Choice-General	95

1980-1989

(2120) Construction,-Analysis,-and-Use-of-Econometric-Models	2193
(2110) Econometric-and-Statistical-Methods-and-Models-General	1718
(1320) Forecasting;-Econometric-Models-General	690
(2113) Distributed-Correlated-Disturbance-Terms;-Inferential-Problems-in-Single-Equation-M...	661
(1324) Forecasting-and-Econometric-Models-Theory-and-Methodology	433
(0230) Macroeconomic-Theory-General	400
(2116) Time-Series-and-Spectral-Analysis	329
(2112) Inferential-Problems-in-Simultaneous-Equation-Systems	279
(1322) General-Forecasts-and-Models	233
(8210) Labor-Economics:-Theory-and-Empirical-Studies-Illustrating-Theory	222

1990-1995/5

(2116) Time-Series-and-Spectral-Analysis	1813
(2120) Construction,-Analysis,-and-Use-of-Econometric-Models	1670
(2114) Econometric-and-Statistical-Methods-and-Models;-Multivariate-Analysis,-Statistical-I...	1654
(2113) Distributed-Correlated-Disturbance-Terms;-Inferential-Problems-in-Single-Equation-M...	1320
(2110) Econometric-and-Statistical-Methods-and-Models-General	1029
(2112) Inferential-Problems-in-Simultaneous-Equation-Systems	569
(1324) Forecasting-and-Econometric-Models-Theory-and-Methodology	330
(3132) Capital-Markets-Empirical-Studies,-Including-Regulation	307
(3131) Capital-Markets:-Theory,-Including-Portfolio-Selection,-and-Empirical-Studies- ...	264
(9211) Consumer-Economics-Living-Standards,-Composition-of-Overall-Expenditures, ...	230

1990-1995/5 sank der Anteil der Publikationen mit einer Nennung der Bayes-Pauschalkategorie auf unter 30 Prozent. Wurden 1980-1989 noch die zweit- bis fünfhäufigsten Kategorien zusammengekommen nicht so häufig genannt wie die erste, so sind es 1990-1995/5 nur noch die zweit- und dritthäufigste, die zusammen etwa so häufig erscheinen. Auffallend ist das ebenso starke Anwachsen von Nennungen der Kategorie 'Time-Series-and-Spectral-Analysis': von Rang 28 (2 Arbeiten) 1969-1979 stieg die Kategorie über Rang 8 (37 Arbeiten) auf Rang 2 (106 Arbeiten) 1990-1995/5. Auch hier ist diese Angabe nach 'Bayesian-Statistics-and-Bayesian-Econometrics' die häufigste Nennung.

Wenden wir uns nun kurz der qualitativen Entwicklung zu. Die bayesianische Behandlung ökonomischer Modelle setzte in den sechziger Jahren mit einer Analyse simultaner Gleichungssysteme ein.³¹¹ Eine große Schwierigkeit bestand zunächst darin, daß die Integration der A-posteriori-Dichte einen erheblichen

Bayesianische Publikationen

1969-1979

(2115) Bayesian-Statistics-and-Bayesian-Econometrics	145
(2113) Distributed-Correlated-Disturbance-Terms;-Inferential-Problems-in-Single-Equation-Models	22
(5110) Organization-and-Decision-Theory	12
(0222) Microeconomic-Theory-Theory-of-the-Household-Consumer-Demand	11
(2112) Inferential-Problems-in-Simultaneous-Equation-Systems	11
(2120) Construction,-Analysis,-and-Use-of-Econometric-Models	9
(0223) Microeconomics-Theory-of-Production	6
(1324) Forecasting-and-Econometric-Models-Theory-and-Methodology	6
(2110) Econometric-and-Statistical-Methods-and-Models-General	6
(0260) Economics-of-Uncertainty-and-Information;-Game-Theory-and-Bargaining-Theory:-General	5

1980-1989

(2115) Bayesian-Statistics-and-Bayesian-Econometrics	320
(2110) Econometric-and-Statistical-Methods-and-Models-General	82
(2120) Construction,-Analysis,-and-Use-of-Econometric-Models	76
(0260) Economics-of-Uncertainty-and-Information;-Game-Theory-and-Bargaining-Theory:-General	68
(0261) Theory-of-Uncertainty-and-Information	50
(0220) Microeconomic-Theory-General	39
(2113) Distributed-Correlated-Disturbance-Terms;-Inferential-Problems-in-Single-Equation-Models	38
(2116) Time-Series-and-Spectral-Analysis	37
(5110) Organization-and-Decision-Theory	19
(2112) Inferential-Problems-in-Simultaneous-Equation-Systems	18

1990-1995/5

(2115) Bayesian-Statistics-and-Bayesian-Econometrics	195
(2116) Time-Series-and-Spectral-Analysis	106
(2113) Distributed-Correlated-Disturbance-Terms;-Inferential-Problems-in-Single-Equation-Models	97
(2114) Econometric-and-Statistical-Methods-and-Models;-Multivariate-Analysis,-Statistical-I...	94
(0261) Theory-of-Uncertainty-and-Information	89
(2120) Construction,-Analysis,-and-Use-of-Econometric-Models	76
(0262) Game-Theory-and-Bargaining-Theory	73
(5110) Organization-and-Decision-Theory	46
(2110) Econometric-and-Statistical-Methods-and-Models-General	44
(0228) Microeconomics-Agent-Theory	40

rechnerischen Aufwand bedeutete, so daß zunächst A-priori-Dichten gesucht wurden, die – mit der Likelihoodfunktion kombiniert – eine A-posteriori-Dichte ergaben, deren Eigenschaften bekannt waren.³¹² Dies führte in der Regel jedoch nicht zu zufriedenstellenden Lösungen. In den siebziger Jahren war durch die Entwicklung leistungsfähigerer Computer und Software die numerische Integration der A-posteriori-Dichten unter allgemeineren Bedingungen bis zu einer Dimension von drei Gleichungen möglich. Erst mit der bahnbrechenden Arbeit von Kloek/van Dijk (1978) konnten A-posteriori-Dichten weitaus komplexerer Modelle berechnet werden.³¹³ Diese seither weiterentwickelten Verfahren wurden dann auf Situatio-

³¹¹ Siehe Drèze (1962), Zellner (1971, S. 248-290), Chetty (1975), Rothenberg (1975), Drèze/Richard (1983).

³¹² Vgl. Bauwens/Lubrano (1993, S. 19), zu den Begriffen oben, Abschnitt A 3 f.

nen abhängiger Zufallsvariablen erweitert und bilden gegenwärtig ein Forschungsgebiet, das auch unter Nicht-Bayesianern starkes Interesse findet.³¹⁴

Auch auf anderen Gebieten wurden bayesianische Alternativen entwickelt. Die Verbreitung bayesianischer Konzepte und Techniken in der Ökonometrie ist zu einem großen Teil A. Zellner – sowohl als Autor als auch als Herausgeber – zu verdanken.³¹⁵ Bei Zellner finden sich auch ausführliche Gedanken zu einer epistemologischen Fundierung der bayesianischen Techniken, wobei seine Konzeption vor allem auf der Theorie von H. Jeffreys beruht.³¹⁶

Einen anderen Ansatz, den man vielleicht am ehesten als vollständige Methodologie oder methodologisches Konzept im Rahmen bayesianischer Methoden bezeichnen könnte, ist die Position von E. E. Leamer.³¹⁷ Betrachten wir zunächst seine Kritik an der Position von Sims.

Leamers Kritik an vektorautoregressiven Modellen bezieht sich, wie schon die Kritik von T. Cooley und S. LeRoy, insbesondere auf die Behauptung, daß kausale Inferenzaussagen in diesem Rahmen möglich seien. Leamer akzeptierte den potentiellen Nutzen vektorautoregressiver Modelle sowohl für Prognosezwecke als auch für eine Deskription vorhandener Daten. Falls eine ökonomische Theorie vorhanden sei, ließe sich ein vektorautoregressives Modell sogar im Rahmen des *Cowles-Commission*-Programms strukturell interpretieren.³¹⁸ Die entscheidende Frage war für ihn in diesem Zusammenhang jedoch die Aussagefähigkeit solcher Modelle bezüglich kausaler Inferenz: "Can causal inferences be made in the context of vector autoregressions without relying on a priori theory? The answer is quite clearly no."³¹⁹ Er verband damit eine grundsätzliche Skepsis, inwieweit sich überhaupt kausale Inferenzaussagen anhand makroökonomischer Daten treffen lassen. Der von Granger vorgeschlagene technische Kausalitätsbegriff sollte durch einen Begriff wie "precedence" ersetzt werden, da der Ausdruck 'Kausalität' inhaltlich vorbelastet sei.³²⁰ Man könne z. B. durchaus annehmen, daß zwischen der Wettervorhersage und dem Wetter ein Zusammenhang besteht, aber niemand werde annehmen, daß die Wettervorhersage das Wetter kausal bedinge. Die bereits

³¹³ Bauwens/Lubrano (1993, S. 19) weisen darauf hin, daß dieser Durchbruch bemerkenswerterweise nicht von Statistikern, sondern von Ökonometrikern erzielt wurde.

³¹⁴ Dies gilt insbesondere für die Markov-Ketten-Simulation und den sog. Gibbs-Sampler.

³¹⁵ Siehe etwa Zellner (1971), Fienberg/Zellner (1975), Zellner (1980), Zellner (1985), Zellner (1988) oder Zellner (1990).

³¹⁶ Siehe oben, Abschnitt A 3 f.

³¹⁷ Siehe vor allem Leamer (1978). Eine Reihe von Schriften wurde in Leamer (1994a) wiederabgedruckt. Aktuelle Informationen zu seinen Forschungstätigkeiten sind im Internet auf einem Server der *University of California* unter der Adresse <http://www.stat.ucla.edu/~leamer> zu finden.

³¹⁸ Vgl. Leamer (1985, S. 256). Wir beschränken uns im folgenden auf eine Auswahl der Argumente Leamers, die für unseren Zusammenhang relevant sind.

³¹⁹ Ebda.

³²⁰ Granger selbst sah dies als unproblematisch an: "Provided I define what I personally mean by causation, I can use the term. I could, if I wish, replace the word cause throughout my lecture by some other words, such as 'oshkosh' or 'snerd', but what would be gained? It is like saying that whenever I use x, you would prefer me to use z." Granger (1980, S. 333).

angesprochene gewisse Willkür, die auch der Variablenauswahl in VAR-Modellen eigen ist,³²¹ war für Leamer ein Hauptkritikpunkt:

“Are the other variables selected at random? Is their choice less whimsical than the choice of instruments in the Cowles-Programm? Should we control for money in the weather-forecasting equation? ‘No,’ probably not,’ and ‘of course not’ are the answers. The variables that we select are based on a priori notions about causal structure. It is thought that the weatherman passively collects information useful for the prediction, and money is rather unlikely to be one of the variables. The point is that I am trying to make here is that a theoretical structure is required to direct the selection of the variables. It is clearly a matter for theoretical speculation whether money is more like a weather forecast or more like a low-pressure system to the west, the latter but not the former relationship being causal. *No matter how diligently you study the time series on money and income, you cannot answer that question.*”³²²

Insbesondere für die Auswahl geeigneter Alternativen (z. B. hinsichtlich der Auswahl instrumenteller Variablen) stelle die klassische Statistik keine Möglichkeit bereit; lediglich ein bayesianischer Ansatz sei geeignet, dieses Problem zu berücksichtigen. Zwar sei damit die Notwendigkeit verbunden, eine A-priori-Verteilung exakt festzulegen, doch könne dieses Problem durch eine angemessene “Sensitivitätsanalyse” hinreichend abgeschwächt werden. Den Grund für den Erfolg vektorautoregressiver Modelle vermutete Leamer weniger in deren empirischen Ergebnissen, sondern in den “Versprechungen”, die mit der Anwendung dieser Modelle gemacht werden: lediglich anhand gegebener Daten (ohne Hinzunahme einer Theorie) über Kausalität und Exogenität entscheiden zu können.³²³ Seiner Meinung nach waren insbesondere die Begriffe “exogen”, “strukturell” und “kausal”, über deren Bedeutung sich Ökonomen durchaus im klaren seien, in der Ökonometrie bislang noch nicht adäquat definiert worden.³²⁴ Er betonte, daß im Zusammenhang mit nichtexperimentellen Daten eine Ursachenforschung, wie sie etwa in Experimenten durch Randomisierung vorgenommen werde, lediglich als Metapher aufgefaßt werden könne:

“Though the econometric literature is replete with procedures to deal with inadequacies in the design, such as simultaneity and selectivity bias, all of these methods ultimately make use of the randomization metaphor and therefore do not require new concepts of statistical inference. For example, phrases such as ‘sampling distributions’ that have a relatively clear meaning when experiments can be repeated are used also in contexts such as the analysis of macroeconomic data in which the notion of repeating an experiment stretches the imagination. In such nonexperimental contexts it is probably better to reject as inappropriate the frequency interpretation of probability and to adopt instead the personal or Bayesian viewpoint in which the metaphorical nature of probabilities is made more or less explicit.”³²⁵

³²¹ Siehe oben, Abschnitt b.

³²² Leamer (1985, S. 286). Hervorhebung von uns.

³²³ Dieses Phänomen sah er als weiteren Beleg für die Thesen McCloskeys über die Rhetorik der Ökonomie, hier McCloskey (1983).

³²⁴ Vgl. Leamer (1985, S. 258). So werde z. B. der Begriff der Exogenität bislang von Ökonometrikern und Zeitreihenanalytikern unterschiedlich verstanden.

³²⁵ Ebda., S. 269.

Leamers Kritik beschränkte sich nicht auf den vektorautoregressiven Ansatz von C. Sims. In einer zwei Jahre zuvor geäußerten, vielbeachteten Grundsatzkritik forderte er: "Let's Take the Con out of Econometrics".³²⁶ Leamer warf auch hier der ökonometrischen Zunft vor, daß sie sich in ihren Handlungen an den experimentellen Wissenschaften orientiere, jedoch nicht über deren Voraussetzungen verfüge. Er illustrierte den Unterschied in einem plastischen Bild, durch das wir auch einen Eindruck von Leamers Stil vermittelt bekommen:

"The applied econometrician is like a farmer who notices that the yield is somewhat higher under trees where birds roost, and he uses this as evidence that bird droppings increase yields. However, when he presents this finding at the annual meeting of the American Ecological Association, another farmer in the audience objects that he uses the same data but came up with the conclusion that moderate amounts of shade increase yields. A bright chap in the back of the room then observes that these two hypotheses are indistinguishable, given the available data. He mentions the phrase 'identification problem,' which, though no one knows quite what he means, is said with such authority that he is totally convincing."³²⁷

Dieser Punkt ist für Leamer zentral. In nichtexperimentellen Wissenschaften wie der Ökonometrie sei es *prinzipiell* unmöglich, Ursachen zu identifizieren. Wenn nur ein Faktor als verursachend angesehen werde, tatsächlich jedoch zwei Faktoren einen systematischen Einfluß ausübten, so könne dies bei alleiniger Betrachtung eines Faktors, auch bei noch so umfangreicher Datenbasis, nicht festgestellt werden. Die Auswirkung dieser Fehlspezifikation sei nicht abzuschätzen: "The misspecification [...] is therefore a pure prior concept. One must decide independent of the data how good the nonexperiment is."³²⁸ Die Ökonomie habe von der Naturwissenschaft den "Mythos" der Existenz einer objektiven wissenschaftlichen Inferenz übernommen. Vor allem in den Lehrbüchern werde davon ausgegangen, daß eine Inferenz objektiv sein könne, da die (Wirtschafts-) Theorie Modellform und Auswahl der Variablen vorgäbe. In der praktizierten Forschung würden jedoch eine Unzahl von Modellformen und Variablenkombinationen mit Computern untersucht und anschließend ein einziges Modell präsentiert:

"This searching for a model is often well intentioned, but there can be no doubt that such a specification search invalidates traditional theories of inference. The concepts of unbiasedness, consistency, efficiency, maximum-likelihood estimation, in fact, all the concepts of traditional theory, utterly lose their meaning by the time an applied researcher pulls from the bramble of computer output the one thorn of a model he likes best, the one he chooses to portray as a rose. [...] Like elaborately plumed birds who have long since lost the ability to procreate but not the desire, we preen and strut and display our *t*-values."³²⁹

Leamer äußerte jedoch nicht nur eine Grundsatzkritik, sondern setzte den kritisierten Vorgehensweisen auch eine Alternative entgegen, die er umfassend erstmals

³²⁶ Leamer (1983).

³²⁷ Ebda., S. 31.

³²⁸ Ebda., S. 33. Siehe hierzu z. B. Maddala (1992, Kap. 4.9).

³²⁹ Leamer (1983, S. 37). Eine ausführliche Darlegung der Wahrscheinlichkeitsauffassung von Leamer, die sich vor allem auf B. de Finetti bezieht (siehe oben, Abschnitt A 3 f), findet sich in Leamer (1987).

1978 beschrieb und später anhand einiger Beispiele illustrierte.³³⁰ Dieser Ansatz läßt sich nach Pagan durch vier Schritte charakterisieren:³³¹

1. Formulierung einer (möglichst allgemeinen) Familie von Modellen.
2. Entscheidung, welche Parameter von Interesse sind und vorläufige Konstruktion von A-priori-Verteilungen für diese Parameter, die die nicht aus den Daten stammende Information wiedergeben.
3. Durchführung einer "Sensitivitätsanalyse", die die Auswirkungen einer bestimmten Wahl von A-priori-Verteilungen auf die Inferenz untersucht. Dies soll durch eine "extreme bound analysis" geschehen. Dazu werden die potentiell in Frage kommenden Variablen in "sichere" und "zweifelhafte" aufgeteilt, wobei das Interesse an der Auswirkung der "sicheren" Variablen besteht. Die übliche Vorgehensweise, alle möglichen Kombinationen mit den "zweifelhafte" Variablen durchzurechnen und die aus eigener Sicht vorteilhafteste zu präsentieren, soll durch diese Analyse ersetzt werden. Die "extreme bounds" für die Koeffizienten ergeben sich demnach, wenn man die A-priori-Annahmen (A-priori-Verteilungen) variiert und von allen denkbaren Variablenkombinationen und A-priori-Verteilungen jeweils den höchsten und niedrigsten resultierenden A-posteriori-Parameterwert betrachtet. In dieser Spannweite ist dann der "wahre" Parameter unabhängig von der Modellspezifikation zu finden. Wie groß diese Spannweiten bei unterschiedlichen Modellspezifikationen konkret werden können, demonstrierte Leamer unter anderem anhand möglicher erklärender Variablen für die Kapitalverbrechensrate amerikanischer Bundesstaaten, die Determinanten der Inflation oder der US-amerikanischen Arbeitslosenrate.³³² Für diesen Schritt verwendet Leamer ein Computerprogramm SEARCH (*Seeking Extreme and Average Regression Coefficient Hypotheses*), betonte jedoch noch 1994, daß dessen Möglichkeiten begrenzt sind: "we do not currently have computer routines that can carry out extensive sensitivity analysis."³³³
4. Die im 3. Schritt ermittelten Spannweiten sollten (z. B. durch Restriktionen) verringert werden. Ist dies unmöglich, sei die anhand der Daten zu erzielende Inferenz "fragil".

Dieser angedeutete Ansatz ist sicher nicht als typisch bayesianisch zu bezeichnen. Leamer hat zwar in vielerlei Hinsicht anregend gewirkt, seine Methodologie aber bislang kaum Nachfolger gefunden.³³⁴

³³⁰ Vgl. Leamer (1978). Chatfield (1995, S. 422) nannte dies "a book sadly neglected by statisticians" und sah darin eine der wenigen Ausnahmen statistischer Arbeiten über spezielle Probleme der Inferenz, die sich explizit mit der Modellsuche auseinandersetzten. Für weitere Aspekte siehe die in Leamer (1994a) wiederabgedruckten Arbeiten.

³³¹ Vgl. Pagan (1987 [1995, S. 15]).

³³² Vgl. Leamer (1983), Leamer (1986), Leamer (1991), zu den technischen Einzelheiten Leamer/Leonard (1983).

³³³ Leamer (1994b, S. xi).

³³⁴ Vgl. Pagan (1995, S. 35).

To Stretch or to Squeeze? D. F. Hendry's Prokrustes-Ökonometrie

Die letzte ihrem Anspruch nach vollständige Methodologie, die wir hier erwähnen wollen, verdanken wir D. F. Hendry.³³⁵ Im Gegensatz zur klassischen *Cowles-Commission*-Ökonometrie geht dieser Ansatz aus einer angelsächsischen Tradition hervor, die von vornherein stärker den Zeitreihencharakter ökonomischer Daten berücksichtigt.³³⁶ Hendry ist der Ansicht, daß das ökonomische Geschehen kompliziert, dynamisch, nichtlinear sowie hochdimensional ist und sich permanent weiterentwickelt. Außerdem werde es von der Entwicklung sozialer Normen und Gesetze sowie durch die technologische Entwicklung beeinflusst. Die dieses Geschehen abbildenden und uns zur Verfügung stehenden Daten – Zeitreihen – seien in der Regel kurz, hochaggregiert, heterogen, nichtstationär, zeitabhängig und interdependent, nicht optimal gemessen oder wandelnden Messungen unterworfen. Auch ökonomischen Theorien, zu deren Überprüfung diese Daten herangezogen werden, durchlaufen Entwicklungen und Revisionen. Dennoch hält Hendry an der Existenz von invarianten Systemen, einem "datengenerierenden Prozeß", fest. Dessen Erkennen sei zwar nicht einfach, aber zweifelsohne möglich.³³⁷ Er hat für eine erfolgreiche Vorgehensweise vier "goldene Regeln" formuliert, die in jeder empirischen Untersuchung befolgt werden sollten:³³⁸

1. *"think brilliantly"*. Jede Analyse sollte die richtigen Antworten möglichst kennen. In diesem Fall seien die empirischen Resultate optimal und bestätigten die eigene Brillanz.
2. *"be infinitely creative"*. Wenn man nicht vor der Analyse ein korrektes Modell unterstelle, sollte man dies während des Konstruktionsprozesses tun. Da keine einheitliche Konstruktionsmethode existiere, könnten die untersuchten Daten die Modellentwicklung auf systematische Weise unterstützen.
3. *"be outstandingly lucky"*. Wenn man weder vor Beginn der Analyse noch während des Konstruktionsprozesses ein 'wahres' Modell unterstelle, werde man früher oder später durch die Anwendung der empfohlenen Methodologie dennoch darüber "stolpern". Falls dies nicht der Fall sein sollte:
4. *"stick to doing theory"*.

³³⁵ Aktuelle Informationen zu seinen Forschungstätigkeiten sind im Internet auf einem Server der *Oxford University* unter der Adresse <http://hicks.nuff.ox.ac.uk/brant/hendry.htm> zu finden. Einen guten Zugang zu Hendrys Ansatz bietet eine Sammlung seiner methodologischen Schriften sowie die vor kurzem erschienene umfassende Monographie, siehe Hendry (1993), Hendry (1995). Der von Hendry vertretenen Richtung gehören unter anderen auch A. Banerjee, M. P. Clements, J. E. H. Davidson, J. A. Doornik, R. Engle, N. R. Ericsson, J.-P. Florens, G. M. Mizon, M. S. Morgan, A. J. Neale, D. Qin, J.-F. Richard, A. Spanos und F. Srba an.

³³⁶ Vgl. zu dieser Entwicklung Gilbert (1989).

³³⁷ Vgl. Hendry (1995, S. 5). Siehe auch Hendry in Hendry/Leamer/Poirier (1990, S. 188, 195). Variable Parameter sind in diesem Zusammenhang konstante Parameter auf höherer Ebene. Für weitere Hinweise siehe Keuzenkamp (1995, S. 236).

³³⁸ Diese Regeln finden sich in Hendry (1987) und in Hendry (1995, S. 5) wiederholt.

Es könnten keine realistischen hinreichenden Bedingungen dafür angegeben werden, daß der Modellkonstruktionsprozeß zu einem "guten" Modell führe, dennoch liefere seine Methodologie einige notwendige Bedingungen, die eine Reihe von schlechten Modellen ausschließe und die Konzentration auf die verbleibenden lenke.³³⁹

Wie auch Leamers Methodologie besteht nach Pagan die konkrete Umsetzung des von Hendry selbst als "Prokrustes-Ökonometrie" bezeichneten Ansatzes³⁴⁰ aus vier Schritten:³⁴¹

1. Zu Beginn sollte ein möglichst allgemeines Modell formuliert werden, in das alle Variablen, die eine Wirtschaftstheorie vorsieht, in möglichst allgemeiner dynamischer Form eingehen.³⁴²
2. In einem zweiten Schritt sollte dieses Modell so reparametrisiert werden, daß die Regressoren annähernd orthogonal und strukturell interpretierbar sind.³⁴³
3. Anschließend sollte das Modell in einer *general-to-specific*-Strategie "in Einklang mit den Daten" so weit wie möglich vereinfacht werden.
4. Das resultierende Modell wird schließlich einer umfassenden Modellkritik ausgesetzt.³⁴⁴ Wie Leamer stützt sich auch Hendrys Methodologie auf ein eigens entwickeltes Computerprogramm, in diesem Falle PC-GIVE. Allerdings ist Hendrys Programm weit umfassender und darüber hinaus kommerziell verfügbar.

Grundlegend für Hendry ist die Vorstellung eines "datengenerierenden Prozesses", in dem feste Parameter existieren. Wie die Abfolge der Schritte von einem allgemeinen zu einem spezifischen Modell in der konkreten Forschungspraxis vor sich geht, bleibt jedoch im Dunkeln.³⁴⁵

Wir wollen hier nicht auf die – mittlerweile zahlreichen – empirischen Untersuchungen eingehen, in denen Hendry seinen Ansatz umsetzte,³⁴⁶ sondern statt

³³⁹ Vgl. Hendry (1995, S. 5).

³⁴⁰ Siehe Hendry/Mizon (1985).

³⁴¹ Vgl. Pagan (1987 [1995, S. 10]).

³⁴² Diese Forderung steht mit dem Ansatz von Sims in Einklang (s. o.). In technischer Hinsicht handelt es sich hierbei zumeist um eine allgemeine dynamische Gleichung oder ein Gleichungssystem, in dem – im Gegensatz zu Sims' vektorautoregressiven Modellen – auch simultane Beziehungen möglich sind. Wie Sims hat auch Hendry in den letzten Jahren seinen Ansatz im Hinblick auf mögliche "Kointegrationen" erweitert.

³⁴³ Anstelle einer solchen Interpretierbarkeitsforderung setzte Sims – wie erwähnt – eine Impulsantwortanalyse.

³⁴⁴ Auch in diesem Zusammenhang hat Hendry mehrfach drei "goldene Regeln" ökonometrischer Forschung formuliert: "Testen, testen, testen." Vgl. Keuzenkamp (1995, S. 234). Damit ist nicht das herkömmliche Testen von Theorien gemeint, wie Poirier (1994b, S. xii) feststellt: "When philosophers of science talk about 'testing' theories they do not have in mind the diagnostic testing that characterizes the British school of econometrics [...] and its controversial attitude towards specification testing."

³⁴⁵ "[...] Hendry [...] normally only provides the value of a test statistic comparing the two models at each end of the path, with very little discussion (if any) of the route followed from one end to the other." Pagan (1987 [1995, S. 13]).

dessen lediglich ein Beispiel anführen, das die berühmte Studie von M. Friedman und A. Schwartz, *Monetary Trends in the United States and the United Kingdom*, aufgriff und mit eben dieser Methodologie reanalysierte.³⁴⁷ Zuvor wollen wir jedoch kurz die für diesen Zusammenhang wesentlichen Elemente des Ansatzes von M. Friedman vorstellen.

Eine pragmatische Alternative: M. Friedman

M. Friedman, der "bekannteste lebende Ökonom der Welt"³⁴⁸, gilt als Begründer des Monetarismus und vehementer Widersacher der keynesianischen Wirtschaftstheorie. Anders als Keynes hat Friedman seine Argumentation jedoch immer wieder mit historischen Untersuchungen gestützt, die in der Tradition des *National Bureau of Economic Research* stehen.³⁴⁹ In methodologischer Hinsicht hatte Friedman bereits 1953 in einem berühmten Aufsatz einen instrumentalistischen Standpunkt dargelegt, und auch seine empirischen Untersuchungen zeigen deutlich diese Einstellung.³⁵⁰

In *Monetary Trends in the United States and the United Kingdom* untersuchten M. Friedman und A. Schwartz die strukturelle Entwicklung monetärer Größen (Geldmengen, Einkommen, Preise und Zinsen) für diese beiden Staaten für den Zeitraum von 1867 bis 1975. Wie schon 1946 A. Burns und W. C. Mitchell legten sie dabei nicht Jahresdaten zugrunde, sondern basierten ihre Analyse auf den Durchschnittsn der vom *NBER* festgestellten Referenzzyklen³⁵¹ und bedienten sich umfangreicher Korrelations- und Regressionsanalysen, ohne die technische Entwicklung der Ökonometrie der vorangegangenen fünfzig Jahre zu berücksichtigen.

Von den mehreren hundert durchgeführten Regressionen wurde von Hendry und Ericsson ein Modell zu Illustrationszwecken reanalysiert: die britische Geldnachfrage von 1878 bis 1970. Für diese Untersuchung wurden

<i>M</i> : die Geldmenge,	<i>RL</i> : die langfristigen Nominalzinsen,
<i>I</i> : das reale Nettoeinkommen,	<i>N</i> : die Bevölkerungsgröße,
<i>P</i> : das Preisniveau,	<i>H</i> : das "high powered money" sowie
<i>RS</i> : die kurzfristigen Nominalzinsen,	<i>P</i> [*] : das Preisniveau der Vereinigten Staaten

herangezogen. Die hier wiedergegebenen Abbildungen geben die Logarithmen der Umlaufgeschwindigkeit v_t und die kurzfristigen Nominalzinsen RS_t (Abb. 17), den

³⁴⁶ Vgl. z. B. die Hinweise in Hendry (1995, passim).

³⁴⁷ Friedman/Schwartz (1982), Hendry/Ericsson (1991).

³⁴⁸ DIE ZEIT Nr. 47, 13.11.1992. Milton Friedman [1912-] promovierte in Chicago und trat dort 1946 eine Professur an. Er erhielt 1980 den Nobelpreis für Wirtschaftswissenschaften.

³⁴⁹ Neben der bereits erwähnten monumentalen Monographie von 1982 ist hier vor allem der nicht minder umfangreiche Vorläufer, Friedman/Schwartz (1963), zu nennen.

³⁵⁰ Friedman (1953). Siehe dazu auch Frazer/Boland (1983). Pelloni (1987) bezeichnet Friedman als "Neo-Bayesianer". Friedman hatte zwar zu Beginn der fünfziger Jahre mehrfach mit L. J. Savage (s. o., Abschnitt A 3 f) zusammengearbeitet und auf diesen auch einen gewissen Einfluß ausgeübt, wie Lindley (1980, S. 692) hervorhebt, andererseits spielten bayesianische Wahrscheinlichkeitskonzeptionen in Friedmans empirischen Untersuchungen keine Rolle.

³⁵¹ Siehe oben, Abschnitt B 2 d.

Logarithmus der realen Geldmenge ($m-p$), sowie den Logarithmus des realen Nettoeinkommens i , (Abb. 18) und den Logarithmus der nominalen Geldmenge m , sowie den Logarithmus des impliziten Preisdeflators p , (Abb. 19) wieder.³⁵² Für die Umlaufgeschwindigkeit unterschieden Hendry und Ericsson aufgrund von Abb. 17 drei historische Perioden.³⁵³ Von 1871 bis 1914 verliefen logarithmierte Umlaufgeschwindigkeit und kurzfristige Zinsen zyklisch und ohne Trend, in der Zwischenkriegszeit von 1918 bis 1938 fiel die logarithmierte Umlaufgeschwindigkeit und wurde volatiler. Beide Größen fielen auf ihr niedrigstes Niveau um 1947 und stiegen daraufhin wieder bis 1970, die Zinsen allerdings mit stärkeren Fluktuationen. Anschließend fiel die logarithmierte Umlaufgeschwindigkeit von 1971 bis 1975, während die Zinsen stiegen. Die Entwicklung der Komponenten von v , verdeutlicht Abb. 18. Während das logarithmierte Realeinkommen zu Beginn der Zwischenkriegszeit stark fiel, stieg der Logarithmus der realen Geldmenge seit etwa 1920 an und blieb oberhalb des logarithmierten Realeinkommens, wobei der Abstand 1947 am größten war. Danach fiel die logarithmierte Geldmenge kontinuierlich bis etwa 1960 bei gleichzeitigem Anstieg des logarithmierten Realeinkommens. Schließlich kam es um 1971 noch einmal zu einem starken Anstieg der logarithmierten Geldmenge. Somit verdoppelte sich die Umlaufgeschwindigkeit zwischen 1947 und 1970, da die logarithmierte Geldmenge sank, während das logarithmierte Realeinkommen stieg. Abb. 19 schließlich zeigt die Logarithmen der nominalen Geldmenge und des impliziten Preisdeflators. Beide Größen weisen einen weitgehend parallelen Trendverlauf auf, wobei die Geldvariable im Vergleich zur Preisvariable wuchs und 1920 bis 1922 ein starker Abfall in p , mit einem geringeren Abfall von m , korrespondierte. Die Ansprüche, die Hendry und Ericsson an die Analyse dieser Daten stellten, sind nicht gering:

“To represent adequately the century of data, an econometric model of U.K. money demand must encompass such features.”³⁵⁴

Betrachten wir zunächst die Analyse von Friedman und Schwartz. Sie schätzten mit diesen Daten (anhand der Phasendurchschnitte) vier Regressionsgleichungen:³⁵⁵

- (a) die logarithmierte Geldmenge pro Kopf ($m-p-n$) in Abhängigkeit des logarithmierten Pro-Kopf-Einkommens ($i-n$), den kurzfristigen Zinsen (gewichtet mit dem Faktor H/M), der Wachstumsrate von $p+i$ sowie zwei Dummy-Variablen für die Nachkriegsperiode und einen Nachfrage-"shift" für die Periode von 1921 bis 1955,

³⁵² Kleine Buchstaben bezeichnen im folgenden Logarithmen.

³⁵³ Die folgende Darstellung nach Hendry/Ericsson (1991, S. 9f).

³⁵⁴ Ebda., S. 11.

³⁵⁵ Hierbei handelt es sich nicht um ein Gleichungssystem, sondern um vier separate Modelle.

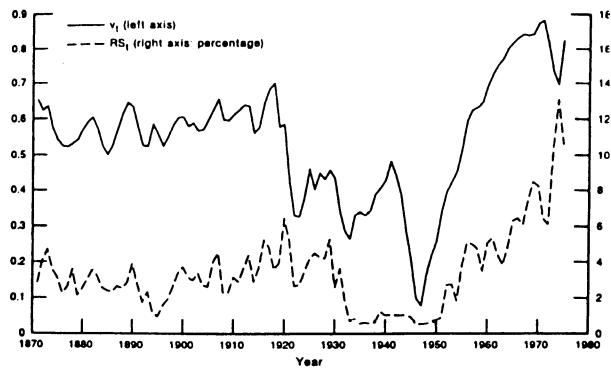


Abb. 16: Logarithmus der Umlaufgeschwindigkeit (v_t) und kurzfristiger Nominalzins (RS_t). Jährliche Daten. Quelle: Hendry/Ericsson (1991, S. 9).

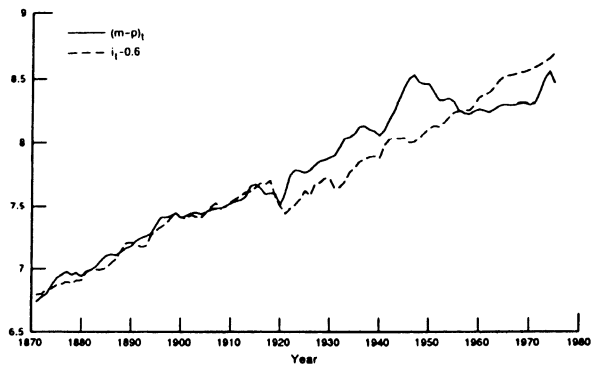


Abb. 17: Logarithmus der realen Geldmenge ($(m-p)_t$) und des Realeinkommens (i_t). Jährliche Daten. Quelle: Hendry/Ericsson (1991, S. 10).

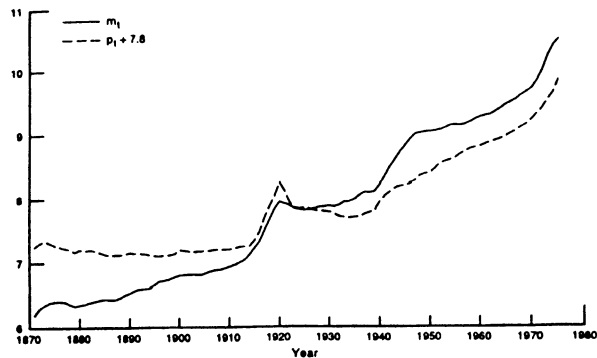


Abb. 18: Logarithmus der nominalen Geldmenge (m_t) und des impliziten Preisdeflators (p_t). Jährliche Daten. Quelle: Hendry/Ericsson (1991, S. 11).

- (b) der Logarithmus des Produkts aus Preisniveau und Nettoeinkommen ($p+i$) in Abhängigkeit von der logarithmierten Geldmenge m , den kurzfristigen Zinsen (gewichtet mit dem Faktor H/M), der Wachstumsrate von $p+i$ sowie zwei Dummy-Variablen ("for postwar adjustment and demand shift"),
- (c) das logarithmierte Preisniveau p in Abhängigkeit von einem linearen Trend, der logarithmierten Geldmenge m , den kurzfristigen Nominalzinsen RS und der Wachstumsrate von $p+i$,
- (d) das logarithmierte reale Nettoeinkommen i in Abhängigkeit von einem linearen Trend, der logarithmierten Geldmenge m , den kurzfristigen Nominalzinsen RS und der Wachstumsrate von $p+i$.³⁵⁶

Falls alle diese Modelle zutreffen sollten, ergeben sich hier, wie Hendry und Ericsson betonten, Widersprüche: Da die Geldmenge in Modell (b), (c) und (d) exogen sei, determiniere Modell (a) das Preisniveau und nicht die Geldmenge; das Preisniveau sei jedoch bereits durch Modell (c) festgelegt.³⁵⁷ Da Friedman und Schwartz (a) als eine ihrer "finalen Gleichungen" auffaßten, legten Hendry und Ericsson diese für ihre weitere Analyse zugrunde und stellten sie einer Reihe von alternativen Modellen gegenüber, wobei sie insbesondere die Probleme der Parameterkonstanz und unberücksichtigter Variablen sowie die Annahmen der Homoskedastizität und Normalverteilung untersuchten. Gleichung (a) lieferte nach Friedman und Schwartz folgende Schätzergebnisse:³⁵⁸

$$(6) \quad (\bar{m} - \bar{p} - \bar{n})_j = \frac{0.16}{(0.08)} + \frac{0.88}{(18.13)} (\bar{i} - \bar{n})_j - \frac{11.16}{(3.42)} RN_j \\ - \frac{0.22}{(0.74)} G(\bar{p} + \bar{i})_j + \frac{1.4}{(2.38)} \bar{W}_j + \frac{21}{(7.56)} \bar{S}_j + \hat{\varepsilon}_j$$

$$J = 36, \hat{\sigma} = 5.54, R^2 = 0.97$$

Die Gleichung liefert die in Abb. 20 wiedergegebene Anpassung. Hendry und Ericsson unterzogen im folgenden dieses Modell einer Reihe von Tests³⁵⁹ und stellten fest, daß es den Daten nicht angemessen ist (nicht alle Parameter sind konstant, die Residuen sind nicht normalverteilt, die Daten weisen einen signifikanten (deterministischen) Trend auf usw.): "Taking this evidence together, equation (1) is not an adequate characterization of the data and is not consistent with the hypothesis of a constant money-demand equation [...]."³⁶⁰

³⁵⁶ Vgl. Friedman/Schwartz (1982, S. 282, 349, 420). Die Schätzgleichungen sind in Hendry/Ericsson (1991, S. 12) wiedergegeben.

³⁵⁷ Hendry und Ericsson beziehen sich hier auf die Exogenitätsdefinitionen von Engle/Hendry/Richard (1983).

³⁵⁸ Vgl. Friedman/Schwartz (1982, S. 282). Die Querbalken markieren, daß die zugrundeliegenden Zeiteinheiten die erwähnten Phasen j sind. Die Werte in Klammern sind t -Werte. Die Schätzung von Hendry/Ericsson (1991, S. 13) weicht hiervon nur unwesentlich ab.

³⁵⁹ Insgesamt unterzogen sie diese Daten (36 Phasen bzw. 93 jährliche Beobachtungen) in ihrer Reanalyse (getreu Hendrys goldenen Regeln) nicht weniger als 46 Teststatistiken, wobei die angegebenen Standardfehler der Parameter nicht mitgezählt wurden. Vgl. Keuzenkamp (1995, S. 242).

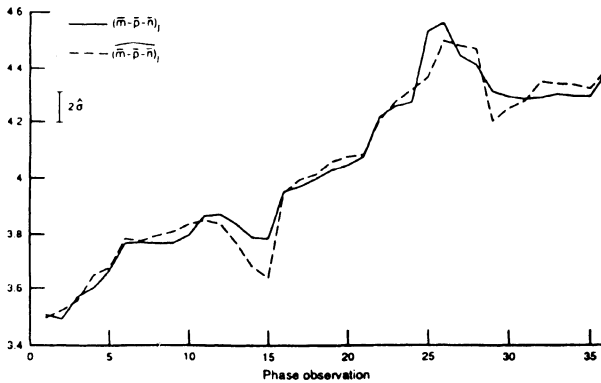


Abb 19: Aktuelle und angepaßte Werte für $(m-p-n)$ nach Gleichung (6). Quelle: Hendry/Ericsson (1991, S. 14).

Anstelle von Modell (6) gingen sie von einem autoregressiven *distributed-lag*-Modell fünfter Ordnung aus, das sie für jährliche Daten von 1878 bis 1970 schätzten und in einer *general-to-specific*-Strategie vereinfachten.³⁶¹ Von den anhand dieses Modells geschätzten 37 Parametern (fünf Variablen sind simultan und mit einer bis fünf Verzögerungen enthalten, zwei weitere mit einer bis drei Verzögerungen) blieben nach der nicht näher erläuterten Vereinfachungsprozedur 10 Parameter in einem Modell mit nahezu vollständig orthogonalen Regressoren, das alle Testkriterien überlebte, übrig.³⁶²

$$(7) \quad \hat{y} = \underset{(0.06)}{0.45} \Delta(m-p)_{t-1} - \underset{(0.04)}{0.10} \Delta^2(m-p)_{t-2} - \underset{(0.04)}{0.60} \Delta p_t + \underset{(0.05)}{0.39} \Delta p_{t-1} - \underset{(0.006)}{0.021} \Delta r s_t \\ - \underset{(0.021)}{0.062} \Delta_2 r l_t - \underset{(0.59)}{2.55} (\hat{u}_{t-1} - 0.2) \hat{u}_{t-1}^2 + \underset{(0.002)}{0.005} + \underset{(0.6)}{3.7} (D_1 + D_3) \quad ,$$

wobei $y_t = \Delta(m-p)_t$.

$$T = 1878-1970, \hat{\sigma} = 1.424, R^2 = 0.87$$

In dieser Form sei das Modell auch ökonomisch interpretierbar. Die Koeffizientenwerte implizieren starke und unmittelbare Reaktionen auf Änderungen der Inflationsrate und der Zinsen, aber insgesamt eine langsame Anpassung durch den Fehlerkorrekturterm.

Diese Reanalyse von Hendry und Ericsson wurde unmittelbar darauf von Friedman und Schwartz kommentiert. Ihre Haltung bezeichneten sie dort im Gegensatz zu

³⁶⁰ Hendry/Ericsson (1991, S. 16).

³⁶¹ Vgl. ebda., S. 22. Die Daten von 1971 bis 1975 wurden nicht verwendet, denn diese "appear to have a different stochastic structure". Ebda.

³⁶² Vgl. ebda., S. 25. Δ bezeichnet Differenzen, u eine Residualgröße aus der Schätzung einer statischen Kointegrationsregression der Gleichung $(m-i-p)_t = -0.309 - 7.00RS_t - u_t$, D_1 und D_3 sind Dummy-Variablen für die Perioden 1914-1918 und 1939-1945, die Werte in Klammern "heteroskedastizität-konsistent" geschätzte Standardfehler.

dem Ansatz von Hendry und Ericsson als "eklektisch": "We believe that there is no magic formula for wringing reasonable conjectures from refractory and inaccurate evidence."³⁶³

Ihrer Ansicht nach eigne sich die Vorgehensweise von Hendry und Ericsson allenfalls zu Demonstrationszwecken für Ökonometrie-Lehrbücher, jedoch nicht, um die Relevanz ihrer Analyse in Frage zu stellen. Auch die Forderung nach formalen Tests beurteilten sie skeptisch:

"[...] we certainly do test for constancy by comparing regressions for different periods, trying to isolate features of the statistical data that may bias comparisons, drawing on historical data, and in a variety of other ways. We regard such tests as far more reliable and, to use a statistical term, robust than the formal statistical tests used by H[endry /] E[ricsson]."³⁶⁴

Darüber hinaus widersprachen sie der von Hendry und Ericsson adaptierten Exogenitätsdefinition. Ob eine Variable als exogen zu bezeichnen sei oder nicht, müsse aufgrund von inhaltlichen und nicht statistischen Kriterien entschieden werden.

Eine weitere Uneinigkeit bestand bezüglich der Interpretation des stochastischen Teils der Gleichungen (des Fehlerterms). Während Hendry und Ericsson offenbar die Fehlergröße als unberücksichtigte Faktoren interpretierten,³⁶⁵ stellte sie für Friedman und Schwartz Fehler in den Variablen (Meßfehler) dar.³⁶⁶ Diese von Friedman und Schwartz als "Regressionseffekt" bezeichnete Größe war für sie erheblich bedeutender als der von Hendry und Ericsson allein berücksichtigte "sampling-error". So ergibt sich in ihrem Modell etwa für den Koeffizienten von (-11.16) bei Berücksichtigung von Fehlern in den Variablen eine *beträchtliche* Spannweite (zweifacher Standardfehler plus Regressionseffekt) von -4.6 bis -46.3. Ein Ansteigen der Zinsrate um einen Prozentpunkt verringert die reale Pro-Kopf-Geldnachfrage um einen Betrag zwischen 4.6 und 46.3 Prozentpunkten. In dieser Spannweite sei die von Hendry und Ericsson gemessene vergleichbare Größe (-7 in der Gleichung in Anm. 364) enthalten. Die von Hendry und Ericsson angeführte geringere Fehlervarianz von Modell (27) gegenüber (26) sei kein Argument für ein besseres Modell, da Hendry und Ericsson eine differenzierte abhängige Variable betrachteten, sie selbst dagegen eine Regression der Niveaus durchgeführt haben.³⁶⁷

³⁶³ Friedman/Schwartz (1991, S. 39).

³⁶⁴ Ebda., S. 41.

³⁶⁵ Eine solche Interpretation ist für die Form der Modelle, die sie betrachten (autoregressive Modelle), zwingend. Vgl. auch Hendry (1995, S. 28).

³⁶⁶ "Fehler in den Variablen" bedeutet hier, daß die Variablen nicht exakt die Größen widerspiegeln, die die ökonomische Theorie vorsieht. Soweit wir sehen, wird diese Fehlerart von Hendry und Ericsson nicht berücksichtigt.

³⁶⁷ Hendry/Ericsson (1991, S. 27) betonten dagegen: "Although [Modell (7)] has a rate of change as the dependent variable, it is an equation in log-levels because of the error correction term". Friedman/Schwartz (1991, S. 46) behaupteten andererseits: "However their 'error correction terms' are useful primarily for short-run analysis of levels, not for the long-run effects that were our sole concern."

Insgesamt sahen Friedman und Schwartz in der Reanalyse von Hendry und Ericsson keine überzeugende Alternative zu ihrem Ansatz. Das einzige Kriterium, das ihrer Ansicht nach ihre Analyse widerlegen kann, ist kein statistischer Test, sondern ein Test des konstruierten Modells mit anderen Daten, d. h. für andere Regionen und frühere oder spätere Perioden. Nur eine solche Überprüfung könne die Validität der Resultate in Frage stellen. Für sie waren zwei Punkte entscheidend: der nicht vorhandene Erkenntnisgewinn gegenüber ihrer eigenen Untersuchung und die logische Inkonsistenz der Vorgehensweise von Hendry und Ericsson:

“After years of experiments, [Hendry und Ericsson] econometric techniques produced a series of models that confirm some of our principal results, contradict none, and are less successful than our equations in terms of their own criterion of variance-dominance. [...] Their estimates are the end result of trying a large number of alternative hypotheses on a single body of data. As a result, it is impossible to specify how many ‘degrees of freedom’ have been used up in the process of reaching the final equations presented, or, put differently, to estimate the probabilities that their results could have arisen from chance.”³⁶⁸

4. Zusammenfassung und kritische Würdigung

Wie schon in Teil A die Konzepte der Statistik wollen wir hier noch einmal die Darstellung der Konzepte der Ökonometrie zusammenfassen und beurteilen. Wir hatten keine systematische Trennung in die Bereiche Deskription, Inferenz und Modellbildung vorgenommen, sondern, um der historischen Entwicklung besser gerecht zu werden, zwei Phasen unterschieden: eine erste Phase, in der die Deskription und Exploration ökonomischer Reihen bzw. Prozesse überwog und eine Phase der Inferenz und Modellbildung.

Die erste Phase war bereits durch eine Übernahme der von Galton und Pearson entwickelten – und in bezug auf Zeitreihen bereits von Hooker übernommenen – Korrelationskonzepte gekennzeichnet. Ein entscheidender Unterschied bestand jedoch: Zwar existierte auch in der Ökonomie ein Theoriegebäude, doch war dies weder einheitlich noch so dezidiert, daß es einer unmittelbaren empirischen Anwendung zugänglich war.³⁶⁹ Insofern dominierte von Anfang an ein explorativer Charakter. Erscheinungen wie “Konjunkturzyklen” waren keine physikalischen Größen, die lediglich gemessen werden mußten, sie waren auch keine biologischen Größen, deren Verteilung beliebig genau ermittelt und deren Einflußfaktoren durch Experimente analysiert werden konnten. Vielmehr waren die Daten 1. passiver Natur und nicht ohne weiteres zu vermehren, sie mußten 2. explizit definiert werden und waren 3. nicht bis in alle Ewigkeit einer gleichen, stabilen Verteilung unterworfen.

³⁶⁸ Friedman/Schwartz (1991, S. 47).

³⁶⁹ Die Wirtschaftstheorien eines L. Walras bis A. Marshall gingen von Gleichgewichtszuständen aus, die sich unabhängig vom historischen Kontext durch immerwährend gleiche Motive menschlichen Handels einstellten. Die ökonomischen Gesetze waren in diesen Theorien zeitlos. Das maßgebliche wirtschaftswissenschaftliche Lehrbuch zu dieser Zeit war Marshalls *Principles of Economics*, das 1920 in 8. Auflage erschien.

Die Verwendung der Korrelationsrechnung war bei Galton theoretisch begründet. Weil die beobachteten Daten z. B. einer bivariaten Normalverteilung entstammten, ließ sich ihre Beziehung zueinander in einem Koeffizienten ausdrücken. Diese Begründung wurde schon bei der ersten Anwendung im sozialwissenschaftlichen Kontext durch Yule aufgegeben.³⁷⁰ Aus Gründen der rechentechnischen Handhabbarkeit wurden die funktionalen Beziehungen als linear angesehen, die Parameter aus dem gleichen Grund mittels des Kleinstquadratverfahrens bestimmt. Yules Autorität – er war einer der führenden Statistiker seiner Zeit – rechtfertigte die Anwendung des biometrischen Instrumentariums, auch wenn die theoretische Begründung dieses Vorgehens zweifelhaft war.

Zwei Aspekte sind in diesem Zusammenhang von besonderer Bedeutung. Erstens: Es bedurfte keiner tiefgehenden statistischen Kenntnisse, um zu erkennen, daß die Struktur sozioökonomischer Phänomene anderer Natur war als die Struktur, die das Wachstum von Pflanzen oder Beziehungen zwischen Körpergrößen von Lebewesen bestimmte. Zweitens: Dies wurde um so deutlicher, als man sich der Analyse von Daten zuwandte, die Zeitreihen darstellten.

Die Analyse des Wirtschaftsgeschehens in seiner Prozeßhaftigkeit fand in der Wirtschaftstheorie keine konkreten Aussagen über Dauer, Form und Beziehungen der Konjunkturzyklen untereinander vor. So gingen die Pioniere empirischer Untersuchungen eigene Wege. Moore, wie schon Jevons, suchte Ersatz in der Astronomie. Nicht nur astronomische Erscheinungen wie die periodisch variierende Anzahl der Sonnenflecken oder die streng periodische Laufbahn der Venus, die sich alle 8 Jahre zwischen Sonne und Erde schob, wurden als Erklärung herangezogen, auch das Instrumentarium der Astronomie, die Periodogrammanalyse, wurde verwendet. Ein solches Verfahren wies den Vorteil auf, daß es "verborgene" Periodizitäten sichtbar machen konnte. Die anfänglich geradezu euphorische Verwendung des Periodogramms wich jedoch bald der ernüchternden Erkenntnis, daß eine wichtige Anwendungsvoraussetzung nicht bestand: die Stabilität des Untersuchungsobjektes.³⁷¹ Konjunkturzyklen waren nicht wie Planetenbahnen von konstanter, mit Meßfehlern meßbarer Dauer, sondern die Länge und Intensität der Zyklen variierte mit der Zeit ebenso wie die Intensität der Störeinflüsse.

Doch damit nicht genug. Ökonomische Daten wiesen in der Regel Trendverläufe auf. Ihre langfristige Entwicklung war somit nicht um stabile Mittelwerte verteilt. In diesen Fällen gab es keine zeitlosen Gleichgewichtszustände, von denen allenfalls transitorische Abweichungen möglich waren, sondern es existierte eine *irreversible* Entwicklung.

Die Lösung dieses Problems bestand jedoch nicht darin, diese Irreversibilität zum Anlaß für eine grundsätzlich andere Sicht zu nehmen, sondern statt dessen wurden zwei Alternativen eingeschlagen: entweder man postulierte auch für diese langfristige Entwicklung einen funktionalen, durch Meßfehler überlagerten Zusammen-

³⁷⁰ Siehe oben, S. 76f.

³⁷¹ Es ist bezeichnend, daß einer der mittlerweile "klassischen" postulierten Konjunkturzyklen, der etwa 40 Monate umfassende "Kuznetszyklus", nicht durch die Periodogrammanalyse, sondern mit eher informellen Mitteln "entdeckt" wurde.

hang in Form eines Polynoms oder einer anderen Trendfunktion (wenn der langfristige Verlauf ein einigermaßen glattes Erscheinungsbild aufwies). Für die Bestimmung des Trends wurde das Kleinstquadratverfahren angewendet, das bereits ein Eigenleben entwickelt hatte und dessen Siegeszug kaum noch aufzuhalten war. Oder man verzichtete ganz auf ein Modell für die langfristige Entwicklung und schloß sie durch die Betrachtung von Abweichungen von einem gleitenden Mittelwert aus. In beiden Fällen war das Ziel jedoch keine umfassende Analyse der (historischen) Entwicklung, sondern eine "Ausschaltung" dessen, was nicht in das Schema zeitlos gleicher Strukturen eingegliedert werden konnte.³⁷²

Insofern ist es naheliegend, daß sich ein Komponentendenken durchsetzte. Voneinander unabhängige Erklärungsfaktoren bestimmten demnach die lang-, mittel- und kurzfristigen Verläufe, wobei die Trendkomponente als genauso störend empfunden wurde, wie die kurzfristige "Rest"komponente. Schwierig war in diesem Zusammenhang die Frage der Korrelation zu beantworten. Die Untersuchung von Trends und Zyklen auf der einen und Korrelationen auf der anderen Seite war kein getrenntes Erkenntnisinteresse, sondern miteinander verwoben. Zur Untersuchung von Korrelationen war nach Aussagen der Statistiker vorher der Trend auszuschalten, und Ziel der Korrelationsanalyse war eine Untersuchung der Übereinstimmung mittelfristiger, d. h. zyklischer Verläufe.³⁷³

Auf der anderen Seite darf aber auch nicht übersehen werden, daß in dieser Formierungsphase die Frage des historischen Wandels ökonomischer Strukturen durchaus problematisiert wurde. Wenn es eine langfristige Trend"komponente" gab, warum sollten dann nicht auch die Beziehungen ökonomischer Größen untereinander langfristigen Veränderungen unterworfen sein? Die Versuche von Hall (1925), Kuznets (1928a), Ezekiel (1928) oder Frisch (1931), die vorhandenen Konzeptionen in Richtung auf zeitabhängige Modelle zu erweitern oder zumindest auf die Unangemessenheit der herkömmlichen Formalisierungen hinzuweisen, waren daher naheliegend.³⁷⁴

Es kann nur spekuliert werden, weshalb dieser Weg nicht weiter verfolgt wurde. Eine mögliche Erklärung, die wir im Zusammenhang mit der Entwicklung der Theorie stochastischer Prozesse erwogen hatten, könnte lauten, daß die technischen Schwierigkeiten in Hinblick auf eine Modellierung zu groß wurden. Da sich aber ohnehin mit diesen Arbeiten kaum inferenzstatistische Implikationen verbanden, erscheint uns eine andere Erklärung plausibler: die vor allem durch die Arbeit von Slutsky 1927 gefundene und kurz darauf von Kuznets einem englischsprachigen Publikum offerierte verblüffende Gleichartigkeit eines ökonomischen konjunkturanzeigenden Index und einer Reihe aufsummierter Zufallsgrößen. Bedeu-

³⁷² Eine der wenigen Ausnahmen, die dem Trendverlauf eine eigenständige Bedeutung beimäßen, war S. Kuznets. Siehe vor allem Kuznets (1930b).

³⁷³ Diese Zusammenhänge wurden nicht immer gesehen. So waren W. Persons' methodologische Vorschläge aus dem Jahr 1919 von dem Ziel motiviert, kurzfristige Prognosen für die Geschäftslage zu entwickeln. Sie wurden aber so allgemein anerkannt, daß Mitchell (1927) mit anderen Zielen (Analyse von Zyklen) diese Methoden als Maßstab empfahl.

³⁷⁴ Vgl. oben, Abschnitt B 2 b.

tete diese Gleichartigkeit, daß auch in der Ökonomie Zyklen lediglich auf Zufallsgrößen zurückzuführen waren?

Wir hatten in Teil A, Abschnitt 4 b dargelegt, wie durch die Untersuchungen von Yule und Slutsky die moderne Theorie stochastischer Prozesse konzeptuell begründet wurde. Beide beschrieben zwar unterschiedliche Modellformen – Yule autoregressive Prozesse und Slutsky sogenannte moving-average-Prozesse –, die Konstruktionen hatten jedoch eine entscheidende Gemeinsamkeit. Sie betrachteten eine Zeitreihe als Realisation einer stochastischen *Differenzengleichung*. Während Yule von einer trigonometrischen Funktion ausging, die man als Differenzengleichung darstellen konnte, in der jedoch der Fehlerterm eine völlig andere Wirkung hatte als in der funktionalen Form, konstruierte Slutsky – auf den ersten Blick recht willkürlich – verschiedene Summen von Zufallsgrößen. Eine tiefere Begründung für die gewählte Modellform (warum etwa eine bestimmte Anzahl von Zufallsgrößen mit verschiedenen Gewichten versehen und ein- oder mehrmals aufsummiert wurde) war dabei von geringerer Bedeutung als die verstörende Tatsache, daß Zufallsgrößen zyklische Erscheinungen hervorrufen konnten.

Es ist erstaunlich, daß offensichtlich bei dieser Konzeptualisierung größere Probleme bestanden, die Vorstellung eines zufälligen, aber dennoch gesetzmäßigen Ablaufes zu akzeptieren, als in der (Querschnitts-) Regressionsrechnung. So wurden Zeitreihen entweder als in den wesentlichen Bestandteilen deterministisch angesehen (Komponentenmodelle) oder als rein zufällig, und dann hatten Zyklen keine Bedeutung. Der entscheidende Punkt wurde übersehen: *Nicht die Zufallsgrößen waren für den (pseudo-) zyklischen Charakter verantwortlich, sondern der Mechanismus, das Modell.*

Diese innere Logik der Modelle blieb – diese These sei hier gewagt – Kuznets genauso verborgen wie später G. Tintner, J. Schumpeter oder J. M. Keynes.³⁷⁵ Es ist daher nicht verwunderlich, wenn Wissenschaftler mit geringerem mathematischen Hintergrund der konzeptionellen Vorstellung, die mit solchen Modellen verbunden war, nicht mehr folgen wollten oder konnten.

Klar erkannt hatte dagegen R. Frisch die innere Logik dieser Modelle und dafür in seinem berühmten *Propagation*-Aufsatz sogar eine ökonomische Rechtfertigung geliefert. In einem dynamischen Modell einer Ökonomie konnten bestimmte Parameterwerte ohne Störeinflüsse gedämpfte Schwingungen hervorrufen, bei der Einwirkung von "Schocks" dagegen die unregelmäßigen Zyklen erzeugen, die erstmals von Yule vorgestellt wurden.³⁷⁶

³⁷⁵ Selbst Tinbergen betonte noch, er "did not understand the role of the shocks as well as Frisch did." Tinbergen in Magnus/Morgan (1987, S. 125).

³⁷⁶ Die Trennung zwischen der Rolle des Mechanismus und der Schocks war für die Entwicklung der Ökonometrie von großer Bedeutung, auch wenn dies Tinbergen im nachhinein kritischer sah: "[...] I think that what interested economics most was not the shocks but the mechanism generating endogenous cycles, and it might very well be that we have overestimated the role of the mechanism. Maybe the shocks were really much more important. This problem was never solved, because the War came along and after the War we were not interested in business cycles anymore." Tinbergen in Magnus/Morgan (1987, S. 125).

Mit der Rezeption dieser Modelle in der Ökonomie scheiden sich die Wege. In dem von Kuznets verfaßten Beitrag "Time Series" in der *Encyclopedia of the Social Sciences* wurde 1934 lediglich das Komponentenmodell ohne stochastische Implikationen beschrieben. Weder Modelle mit variablen Parametern noch die grundsätzliche Bedeutung der Modelle von Yule und Slutsky fanden Erwähnung.³⁷⁷ In diese Richtung gingen dann auch die Arbeiten von Schumpeter sowie von Burns und Mitchell. Schumpeter verfaßte zwar noch den Eröffnungsartikel der 1933 erschienenen ersten Ausgabe der *Econometrica*, nahm aber an der Entwicklung der Ökonometrie keinen Anteil mehr.³⁷⁸

Von entscheidender Bedeutung für die weitere Entwicklung ist die Tatsache, daß die wissenschaftliche Ausrichtung der Ökonometrie maßgeblich von Personen bestimmt wurde, die eine physikalische Ausbildung genossen, wie etwa Jan Tinbergen, Ragnar Frisch, Tjalling Koopmans, Charles Roos oder Harold T. Davies.³⁷⁹ In diesen Köpfen manifestierte sich ein anderes Bild der Ökonomie als das der "traditionellen" empirischen Forscher. Sie übertrugen ein mechanistisches, rigoros mathematisches Modelldenken in die empirische Forschung. Beispielhaft für diese Entwicklung ist eine Schilderung von Koopmans über seinen Werdegang:

"Why did I leave physics at the end of 1933? In the depth of the worldwide economic depression, I felt that the physical sciences were far ahead of the social and economic sciences. What had held me back was the completely different, most verbal, and to me almost indigestible style of writing in the social sciences. Then I learned from a friend that there was a field called mathematical economics, and that Jan Tinbergen, a former student of Paul Ehrenfest, had left physics to devote himself to economics. Tinbergen received me cordially and guided me into the field in his own inimitable way. I moved to Amsterdam, which had a faculty of economics. The transition was not easy. I found that I benefited more from sitting in and listening to discussions of problems of economic policy than from reading the tomes. Also, because of my reading block, I chose problems that, by their nature, or because of the mathematical tools required, have similarity to physics."³⁸⁰

³⁷⁷ Vgl. Kuznets (1934).

³⁷⁸ Schumpeter verfügte über keine große mathematische Begabung. Er selbst schrieb in Hinblick auf die zunehmende Mathematisierung der Ökonomie: "Manchmal fühle ich mich wie sich Moses gefühlt haben muß, als er das Gelobte Land sah und wußte, daß er selbst dorthin nicht gelangen werde." Schumpeter in einem Brief an G. Haberler vom 20.3.1933, zitiert nach Swedberg (1994, S. 164). Schumpeters *Business Cycles* stand in methodischer Hinsicht den Arbeiten von Mitchell und Kuznets viel näher als etwa denen von Frisch oder Koopmans.

³⁷⁹ Siehe Epstein (1987, S. 75 Anm. 39), Mirowski (1989c, S. 234) und vor allem Boumans (1993). Auch der Statistiker G. U. Yule, der sich bei seinen Untersuchungen auf dem Gebiet der Zeitreihenanalyse insbesondere mit Anwendungsmöglichkeiten in der Ökonomie beschäftigte, begann seine wissenschaftliche Karriere mit dem Studium elektrischer Wellen.

³⁸⁰ Zitiert nach Mirowski (1991, S. 152). Hier darf eine soziologische Komponente nicht unterschätzt werden. Es sei die These gewagt, daß durch die starke Mathematisierung, die stets den Nimbus des Wissenschaftlichen hatte, bestimmte, als nicht wissenschaftlich angesehene Forscher ausgeschlossen werden sollten. Für die Mathematisierung der ökonomischen Theorie wurde dieses Argument bereits mehrfach vorgebracht, etwa von A. W. Coats, G. Stigler, J. Maloney oder zuletzt von Porter (1993, S. 63). Frisch und Koopmans führten die Matrizenrechnung, die Mitte der zwanziger Jahre in der Physik Verbreitung fand, im Rahmen der multiplen Regressionsrechnung in die Ökonometrie ein und erschwerten damit eher das Verständnis dieser Texte bei Ökonomen. Vgl. Mirowski (1989c, S. 231).

In diesem Umfeld war 1. eine Modellierung der ökonomischen Welt in Form von Differenzgleichungen und 2. auch eine rigide Stochastifizierung möglich. Auf den ersten Blick erscheint es allerdings merkwürdig, daß Koopmans seinen Ansatz auf die Theorie R. A. Fishers aufbaute und damit nicht, wie noch Frisch, in physikalischer Analogie Meßfehler als Begründung für eine stochastische Betrachtungsweise ansah, sondern in biologischer Analogie von hypothetisch infiniten Populationen ausging, aus denen sich mit konstanten Wahrscheinlichkeiten die vorliegenden Daten realisiert hätten. Vermutlich war hier nicht so sehr die stochastische Grundkonzeption von Bedeutung als vielmehr die Tatsache, daß Fisher eine umfassende statistische Schätztheorie entwickelt hatte³⁸¹ und als führender Statistiker galt.

Die univariate Zeitreihenanalyse entwickelte sich in diesem Zusammenhang zu einem Nebenschauplatz, statt dessen dominierte ein Denken in "vollständigen" Modellen.³⁸² Diese Modelle stimmten jedoch von Beginn an nicht völlig mit den wirtschaftstheoretischen Modellen überein, deren Überprüfung doch das Ausgangsziel der Ökonometrie darstellte. Schon Tinbergen sah sich zu einer Reihe von Kompromissen gezwungen, da die seinerzeit vorhandenen Konjunkturtheorien nicht so spezifiziert waren, daß sie eine unmittelbare empirische Überprüfung ermöglichten.

Die unbefangene, iterative Vorgehensweise Tinbergens verstieß gegen die Regeln der gerade erst von Frisch und Koopmans übernommenen stochastischen Konzeptionen der Statistik. Insofern war die Kritik von Keynes oder Friedman berechtigt. Nichtsdestotrotz wurde der eingeschlagene Weg weiter verfolgt und von einem Schüler R. Frischs, T. Haavelmo, manifestartig fundiert.

Haavelmos Ansatz, der für die weitere Entwicklung richtungweisend wurde, forderte wie derjenige Koopmans eine rigorose stochastische Betrachtungsweise. Anders als Koopmans berief sich Haavelmo jedoch nicht auf die Theorie Fishers, sondern auf diejenige von Neyman und Pearson. Betrachtet man die Grundlagen dieser Theorie, muß ihre Übertragung auf (makro-) ökonomische Entwicklungen problematisch erscheinen.

Wir hatten in Abschnitt A 3 e darauf hingewiesen, daß die Akzeptanz des Neyman-Pearson-Ansatzes eine Konzeption mit sich bringt, die auf Verhaltensregeln abzielt. Schon die Übertragung der Fisherschen Vorstellung hypothetisch infiniten Populationen, aus denen Zufallsstichproben gezogen werden, kann befremdlich erscheinen. Dies gilt jedoch um so mehr für die Neyman-Pearson-Konzeption eines "repeated sampling from the same population". Bei der Übertragung einer solchen Vorstellung auf makroökonomische Zeitreihen ist die Frage zu stellen: "[...] how often is the question that an econometrician has to answer a decision problem in the context of repeated sampling?"³⁸³

³⁸¹ Siehe oben, Abschnitt A 3 d.

³⁸² Zu Beginn der ökonometrischen Entwicklung fanden sich aber durchaus noch Untersuchungen, die in der Tradition der in Abschnitt B 2 b beschriebenen Entwicklung standen. Siehe z. B. Hotelling (1934), Schultz (1934), Greenstein (1935), Regan (1936). Selbst die Methode der gleitenden Mittelwerte wurde hier von Sasuly (1936) noch empfohlen. Siehe auch die Überblicke von Lorenz (1935) und Roos (1936).

Warum legte Haavelmo aber gerade diesen Ansatz zugrunde?³⁸⁴ Eine mögliche Erklärung könnte lauten, daß zu Beginn der vierziger Jahre aus den rivalisierenden Ansätzen von Fisher und Neyman/Pearson letzterer als Sieger hervorgegangen war und im Kuhnschen Sinne bereits ein "Paradigma" darstellte. Ein weiterer Grund ist privater Natur: Haavelmo selbst berichtete, daß er das Privileg genoß, einige Monate bei dem "weltberühmten Statistiker" J. Neyman studieren zu dürfen. Dieser habe ihm, der damals "jung und naiv" gewesen sei, "verheißungsvollere Wege [...] sich dem Problem der ökonometrischen Methodik zu nähern, als jene, die bisher solche Schwierigkeiten und Enttäuschungen verursacht hatten", aufweisen können.³⁸⁵

Haavelmo sah durchaus die Problematik einer einfachen Übertragung der Neyman-Pearson-Konzeption und argumentierte daher instrumentalistisch. Immer wieder finden sich in seinen Ausführungen Redewendungen wie "it has been found fruitful" o. ä. Darüber hinaus basieren große Teile seiner Ausführungen lediglich auf "Hoffnungen":

"[...] we might hope to find elements of invariance in economic life, upon which to establish permanent laws [...]. Our hope in economic theory and research is that it may be possible to establish constant and relatively simple relations [...]. Our hope for simple laws in economics rests upon the assumption that we may proceed as if such natural limitations of the number of relevant factors exist."³⁸⁶

Kann man bei einer solchen Auffassung mit gutem Recht von einer objektiven Inferenz ausgehen? Auch von der problematischen Fundierung einmal abgesehen, lieferte der Neyman-Pearson-Ansatz auf eine Reihe von Fragen keine Antwort. So stellt Heckman zu Recht fest, daß z. B. der wichtige Aspekt der Modellkonstruktion und -selektion von Haavelmo nicht berücksichtigt wurde:

"These claims have never been rigorously established, even for analyses conducted on large samples. There is no 'correct' way to pick an empirical model and the problems of induction, inference, and model selection are very much open. [...] The Neyman-Pearson theory espoused by Haavelmo and the Cowles group takes a narrow view of science. By its rules, hypotheses are constructed in advance of knowledge of the data and the role of empirical work is to test the hypotheses. This rigid separation of model construction and model verification was a cornerstone of classical statistics circa 1944. Even then, influential scholars, primarily Bayesians such as Harold Jeffreys quarreled with this view of empirical science. Since that time, the monopoly of classical statistics has broken."³⁸⁷

³⁸³ Keuzenkamp/Magnus (1995, S. 18).

³⁸⁴ Auch Heckman (1992, S. 881) fragt in diesem Zusammenhang in einer Kritik an Morgan (1990): "Why was the Neyman-Pearson theory adopted as the paradigm of statistical inference in econometrics, and why were rival theories by Ronald Fisher and Harold Jeffreys less successful?"

³⁸⁵ Haavelmo (1994, S. 75).

³⁸⁶ Haavelmo (1944, S. 13, 22f, 24).

³⁸⁷ Heckman (1992, S. 882). Er begründet Morgans Überschätzung des Haavelmo-Ansatzes – unserer Ansicht nach zu Recht – mit ihrer auf den Einfluß Hendry zurückgehenden Ansicht, daß diese Probleme im Rahmen des Neyman-Pearson-Ansatzes überhaupt lösbar sind. Siehe auch Malinvaud (1991, S. 635) und Zellner (1992, S. 220).

Dennoch bestimmte Haavelmos Übernahme des Neyman-Pearson-Paradigmas die ökonometrische Forschung für mehrere Jahrzehnte. Selbst Koopmans berief sich nicht länger auf Fisher, sondern verteidigte gegenüber R. Vining Haavelmos Ansatz. Das physikalische Weltbild war damit zementiert. Koopmans verglich die "vollständigen" Strukturgleichungssysteme mit der Erklärungsmacht der Newtonschen Gravitationstheorie; J. Marshak, Vorsitzender der *Cowles Commission*, verstand seine Aufgabe sogar explizit als "social engineering". erinnert dies nicht in fataler Weise an die – seinerzeit vehement abgelehnte – "Sozialphysik" eines Quetelet?

Im nachhinein erschien selbst den Anhängern dieses Ansatzes das Erreichte gering im Vergleich zu den ursprünglichen Hoffnungen. Die Ökonometrie mußte feststellen, "[...] daß dem großen Trompeten-Solo von Trygve Haavelmo im Jahre 1944 eher kläglicher Chorgesang folgte. Zu einer stochastischen Grundlegung der Ökonometrie ist es leider bis heute nicht gekommen. Die Probleme der Inversion verallgemeinerter Matrizen waren zu drängend."³⁸⁸

Seit den siebziger Jahren wurde daher verstärkt nach alternativen Wegen gesucht. Wir hatten zunächst den Ansatz von C. Sims dargestellt, der den traditionellen simultanen Gleichungssystemen vektorautoregressive Zeitreihenmodelle entgegenstellte. Diese Modelle lieferten zunächst nichts anderes als eine Beschreibung der in den vorliegenden Zeitreihen vorhandenen verzögerten Korrelationsstruktur. Gemäß der sozialwissenschaftlichen Binsenweisheit, daß alles mit allem korreliert (aber nur schwach), könnte man vektorautoregressive Modelle als ideale Modellform ansehen. Mit den Modellen ist jedoch das gleiche Problem wie mit den univariaten ARIMA-Modellen verbunden:³⁸⁹ Das "richtige" Modell muß anhand der Daten erst gefunden werden, was wiederum die Annahmen der klassischen Inferenz verletzt. Darüber hinaus kann man das in der univariaten Zeitreihenanalyse von Box und Jenkins entwickelte Instrumentarium wegen der hohen Komplexität der Modelle nicht verwenden. Sims schlug daher vor, die hohe Anzahl der Parameter, die solche Modelle mit sich bringen, zu restringieren und befürwortete hierfür letztendlich einen bayesianischen Ansatz. Ungeachtet dieser inferenzstatistischen Implikationen sollte allerdings darauf hingewiesen werden, daß für einen konkreten Datensatz dieser Ansatz eine umfassende *Deskription* liefert, die auch – in Abhängigkeit des gewählten Modells – Auskunft über die Auswirkung von externen "Schocks" auf die einzelnen Bestandteile des Systems geben kann. Bayesianische Ansätze hielten in die Ökonometrie in den sechziger Jahren über Strukturgleichungsmodelle Einzug und sahen sich in technischer Hinsicht noch

³⁸⁸ Menges (1980b, S. 163). Seit dem Ende der vierziger Jahre wurde die statistische Verteilungstheorie für Strukturgleichungsmodelle kaum weiterentwickelt, und die nächsten zwei Generationen von Ökonometrikern begnügten sich mit asymptotischen Resultaten, die die *Cowles Commission* lediglich als provisorisch angesehen hatte. Vgl. Epstein (1987, S. 8).

³⁸⁹ Siehe oben, Abschnitt A 4 b. Sie waren auch den gleichen statistischen Beschränkungen wie etwa Stationarität, Linearität unterworfen. Das von Box in DeGroot (1987, S. 249) zur Sprache gebrachte, grundsätzliche Problem, daß "[...] although there's only one way something can be linear, there's an infinite number of ways it can be nonlinear", galt auch für diese allgemeinen Modelle.

größeren Schwierigkeiten ausgesetzt als die klassische Inferenzstatistik. Doch sollten diese technischen Schwierigkeiten nicht darüber hinwegtäuschen, daß der bayesianische Standpunkt aus konzeptioneller Sicht einheitlicher ist:

“That there is a unified and operational approach to problems of inference in econometrics and other areas of science is a fundamental point that should be appreciated. Whether we analyze, for example, time series, regression, or ‘simultaneous equation’ models, the approach and principles will be the same. This stands in contrast to other approaches to inference that involve special techniques and principles for different problems.”³⁹⁰

Am konsequentesten entwickelte E. E. Leamer eine bayesianische ökonomische Methodologie. Die wichtigste Kritik Leamers erscheint uns diejenige bezüglich der Modellproblematik. Leamer betonte zu Recht, daß die klassische Theorie, in der das Modell als gegeben angesehen wird, eine geradezu “Orwellsche” Ökonometrie erfordere:

“In such a fanciful world, personal uncertainties and public disagreements concerning how to interpret data would be completely resolved in advance. New data sets would not be distributed to humans at all, but instead would be delivered with elaborate security measures to a centralized warehouse where preprogrammed computers would pore over the numbers and pass the conclusions to the public. Once analyzed, the data would be entirely destroyed, to prevent the urge to try something else from becoming an unwanted reality.”³⁹¹

Der nichtexperimentelle Charakter der Ökonometrie verbietet eine derartige Vorstellung. Daten wie etwa die Entwicklung des Bruttosozialprodukts eines Landes liegen nur einmal vor, werden jedoch vielfach ausgewertet. Wenn Unsicherheit über das Modell herrscht und bei der Auswahl der relevanten Variablen 1. die Daten nicht unbeteiligt sind und 2. die persönliche Überzeugung des Wissenschaftlers eine Rolle spielt (z. B. Auswahl der Determinanten der Kriminalität durch konservative oder liberale Forscher, Auswahl der Determinanten der Inflation durch Monetaristen oder Keynesianer), dann ist unseres Erachtens ein bayesianischer Standpunkt der einzige, der sich rechtfertigen läßt. Die Angabe der Auswirkung verschiedener Annahmen und Variablenauswahlen, die “Sensitivitätsanalyse”, scheint hier ein vielversprechender Weg zu sein, allerdings müßte eine größere Zahl von Anwendungen in Zukunft deren Leistungsfähigkeit noch untermauern. Die dritte besprochene Methodologie wurde von D. Hendry entwickelt. Anders als Leamer ist Hendry überzeugt, daß eine Modellkonstruktion, die auf der intensiven Analyse eines Datensatzes aufbaut, mit den Mitteln der klassischen Inferenz zu rechtfertigen ist. Wir hatten als Beispiel für seine Vorgehensweise die Reanalyse eines ausgewählten Modells der umfassenden Untersuchung von M. Friedman und

³⁹⁰ Zellner (1971, S. 11). Dieser Punkt wird oftmals ignoriert, so etwa in der umfangreichen Diskussion um die sogenannten Einheitswurzeltests. Sims/Uhlig (1991) betonen ausdrücklich, daß im Rahmen der bayesianischen “Philosophie” die Frage, ob der zugrundeliegende Prozeß stationär ist, unerheblich sei, da bezüglich der Daten lediglich die Likelihoodfunktion interessiere. Darauf hatten bereits Barnard/Jenkins/Winsten (1962) hingewiesen. Siehe dazu unten, Anhang 2 b.

³⁹¹ Leamer (1994b, S. ix).

A. Schwartz über monetäre Trends in England und den Vereinigten Staaten angeführt. An diesem Beispiel wurde ein Problem deutlich, daß A. Pagan prinzipiell konstatierte: Die einzelnen Schritte des Modellbildungsprozesses bleiben im Dunkeln. Die Möglichkeit einer Absicherung anhand von klassischen, auf die Theorie von Neyman und Pearson aufbauenden "Tests" erscheint uns daher höchst problematisch.³⁹²

Ist die Vorgehensweise M. Friedmans, die von ihm selbst nicht als geschlossene Methodologie, sondern eher eklektisch verstanden wird, eine Alternative? Unseres Erachtens ist sie dies nur, wenn man wie schon G. Box (s. o., Abschnitt A 4 c) einen rigoros instrumentalistischen Standpunkt vertritt. Die prinzipielle Beschränkung auf Eingleichungsmodelle, von denen dann allerdings sehr viele betrachtet werden, bedeutet bereits eine erhebliche Reduktion der Komplexität und erweist sich als problematisch, wenn einzelne Gleichungen zueinander im Widerspruch stehen. Friedman³⁹³ vertraut nach eigener Aussage keinen formalen statistischen Kriterien. Er wies schon in seiner Kritik an Tinbergens Überprüfung von Konjunkturtheorien zu Recht darauf hin, daß die klassischen Signifikanz- oder Hypothesentests ihre Aussagekraft verlieren, wenn sie nach einer Datenanalyse für dieselben Daten verwendet werden. Dies hindert ihn jedoch nicht daran, in der eigenen Arbeit regen Gebrauch von *t*-Tests zu machen.

Betrachtet man die Methodologien/Vorgehensweisen insgesamt, so ist trotz Kritik und Weiterentwicklung der klassischen Ökonometrie die Vorstellung eines naturwissenschaftlichen Weltbildes bei Sims und Hendry, in schwächerer Form letztendlich auch bei Leamer und Friedman, dominierend. Vor allem wird grundsätzlich von konstanten, zeitinvarianten Parametern ausgegangen. Die Überprüfung der Parameterkonstanz gehört zwar zum Programm Hendrys, es werden jedoch – abgesehen von Dummy-Variablen – keine Alternativen modelliert. Friedman und Schwartz weisen zwar auf die Bedeutung der Analyse historisch einheitlicher Perioden hin, für diese Perioden unterstellen sie aber wiederum starre Korsette. Bei Leamer und Sims spielen, soweit wir sehen, variable Parameter ebenfalls keine Rolle.³⁹⁴ Die Komplexität wird stets reduziert auf *eine* Parametermatrix, die die Struktur zeitinvariant wiedergibt, gleichgültig ob es sich um kurz- oder langfristige Beziehungen handelt.

Uns erscheint eine solche Vorstellung in bezug auf sozioökonomische Strukturen fragwürdig. Hatte nicht bereits Leibniz in seiner Kritik an Bernoulli von der "Veränderlichkeit der Dinge" gesprochen? Hatte nicht W. Weldon K. Pearson vorgeworfen, daß sich Verteilungen wie die der Intensität des Pauperismus innerhalb von weniger als zehn Jahren veränderten? Waren nicht in den zwanziger Jahren dieses Jahrhunderts Vorstellungen zeitvariabler Parameter durch die nur scheinbar kontingenten, in ihrem Mechanismus jedoch zeitinvarianten Modelle eines Yule

³⁹² Keuzenkamp (1995, S. 243) verwendet daher für Hendrys Vorgehen den treffenderen Ausdruck "diagnostic checks" anstelle von "diagnostic tests".

³⁹³ Wir wollen in diesem konkreten Beispiel davon ausgehen, daß die methodischen Elemente der Untersuchung auf Friedman zurückgehen.

³⁹⁴ Eine in Doan/Litterman/Sims (1984) erwogene Möglichkeit variabler Parameter wurde nicht weiter verfolgt.

oder Slutsky verdrängt worden? Hatte nicht Yule selbst Moore vorgeworfen, daß über einen so langen Zeitraum wie die von ihm untersuchten vierzig Jahre die Struktur der ökonomischen Reihen sich vermutlich verändert hatte? Und hatte nicht sogar der Begründer der modernen häufigkeitstheoretischen Wahrscheinlichkeitsdefinition, J. Venn, die Anwendung wahrscheinlichkeitstheoretischer Aussagen in bezug auf gesellschaftliche Phänomene abgelehnt, eben weil sich derartige Zusammenhänge im Laufe der Zeit änderten?

Diese berechnete Kritik macht es unseres Erachtens nötig, *grundsätzlich* über die Angemessenheit der bisherigen ökonometrischen Konzepte nachzudenken. Wesentliche Aspekte dieses Problems sind unserer Ansicht nach von dem Ingenieurwissenschaftler R. E. Kalman erkannt worden, dessen Kritik wir hier vorstellen wollen.

Kalman hat sich zu Beginn der achtziger Jahre selbst mit dem Problem der Modellkonstruktion in der Ökonometrie befaßt und in diesem Zusammenhang grundlegende Kritik aus systemtheoretischer Sicht geäußert.³⁹⁵ Seiner Ansicht nach verfolgte die Ökonometrie vor allem die beiden folgenden Wege:

1. Ökonomische Gesetze und Beziehungen wurden als dynamische Gleichungen im Sinne Newtonscher Gesetzmäßigkeiten formuliert und
2. die Koeffizienten dieser Gleichungen wurden quantitativ durch die Extraktion statistisch relevanter Information aus realen Daten bestimmt.

Im Hinblick auf diese Entwicklung stellt er fest, daß der dadurch erzielte Erkenntnisfortschritt selbst im Vergleich zu dem 250 Jahre zuvor nach Newton erzielten Erkenntnisfortschritt enttäuschend gering ausgefallen ist. Er vertritt die These (die seiner Ansicht nach in den "harten" Wissenschaften keiner Diskussion bedarf),

"[...] that economics is not at all like physics and therefore that it is not accessible by a methodology that was successful for physics. Far from being governed by absolute, universal, and immutable laws, economic knowledge, unlike physical science, is strongly *system* (context) *dependent*; when economic insights are taken out of temporal, political, social, or geographical context, they become trivial statements with little information content. [...] Since economic 'laws' do not process the attributes of physical laws, writing down equations, in the style of physics, to translate economic statements into mathematics is not a productive enterprise. [...] System theory provides a simple but hard suggestion: Do not write equations expressing assumed relationships; deduce your equations from real data. [...] To put it differently, there will never be a Newton in economics; the path to be followed *must* be different."³⁹⁶

Noch negativer fällt sein Urteil hinsichtlich des zweiten Schrittes, der statistischen Bestimmung von eindeutigen Parametern, aus. Dies sei seiner Ansicht nach nur dann sinnvoll, wenn konkrete und explizit meßbare Parameter vorliegen, wie dies zum Beispiel beim Widerstand im Ohmschen Gesetz der Fall sei:

"Economists have often dreamed of imitating the simple situation characterized by Ohm's law just by hoping for the best, for example, by assuming that such a law (the Phillips curve) exists

³⁹⁵ Wir stützen uns hier vor allem auf Kalman (1982a) und Kalman (1982b).

³⁹⁶ Kalman (1982a, S. 19f). Hervorhebungen im Original.

between inflation and unemployment. But unemployment and inflation, in any quantitative sense, are fuzzy and politically biased attempts to replace complex situations by (meaningless) numbers; consequently any hope that two such concepts can be tied to one another by a single coefficient is barbarously uninformed wishful thinking.”³⁹⁷

Er fordert statt dessen eine rigorose Anwendung der Systemtheorie. Die Systemtheorie geht nicht von einer direkt meßbaren Beziehung zwischen Input und Output aus: “instead of determining a single parameter, such as a resistance, system theory is concerned with the much more general question of determining a system.”³⁹⁸ Parameter in Systemen haben nach Kalman eine völlig andere Bedeutung, als bislang von den Ökonometrikern angenommen; sie sind demnach nur *lokal* zu definieren. Für Kalman ist es keineswegs selbstverständlich, daß das Erkenntnisziel statistischer Analysen in der Erlangung einzelner Zahlen, wie etwa bei der Anwendung der Maximum-Likelihood-Schätzung oder des Kleinstquadratverfahrens, liegt: “[...] common sense should tell us, that such a miracle is possible only if additional assumptions (*deus ex machina*) are imposed on the data which somehow succeed in neutralizing the intrinsic uncertainty.”³⁹⁹ So sei das Kleinstquadratverfahren so beliebt, weil es eine eindeutige (“unique”) Antwort liefere. Die mit einer solchen Vorgehensweise verbundenen Annahmen seien in der Regel jedoch nicht zu rechtfertigen.

Die verbreitete Intention, aus Daten, die eine Streuung aufweisen, einen konkreten Wert zu bestimmen, der die maximale Likelihood aufweise oder Abweichungen minimiere, also allen anderen vorzuziehen sei, ist für ihn “fundamentally wrong and extremely harmful to scientific progress.”⁴⁰⁰ Solch ein Vorgehen impliziere die folgenden Annahmen (“prejudices”):

- (i) Die Daten wurden durch einen probabilistischen Mechanismus generiert.
- (ii) Dieser probabilistische Mechanismus ist sehr einfach; er ist in der Zeit konstant, und eine Verteilungsfunktion erklärt alles.
- (iii) Es gibt einen “wahren” Wert, der als “particularly striking feature” der hypothetischen Verteilungsfunktion angesehen werden kann, wie etwa der Erwartungswert, Median oder Modalwert.

³⁹⁷ Ebda., S. 20. In der Astronomie z. B. bestehen solche eindeutigen Beziehungen, deren Parameter eine direkte system-unabhängige Bedeutung haben, wie die Bestimmung der Position eines Objektes als Funktion des Moments und des Winkels. Vor diesem Hintergrund überrascht es nicht, daß Kalman besonders dem Haavelmo-Ansatz gegenüber kritisch eingestellt ist: “The aspiration of Haavelmo to give a solid foundation to econometrics by dogmatic application of probability theory has not been fulfilled (in the writer’s opinion), no doubt because probability theory has nothing to say about the underlying system-theoretic problems.” Dieser Satz, der eigentlich in Kalman (1982c) erscheinen sollte, wurde dort auf Verlangen eines Gutachters gestrichen und findet sich statt dessen in Kalman (1982b, S. 194).

³⁹⁸ Kalman (1982a, S. 23). Sinnvolle Annahmen für ein solches System seien Linearität und Endlichkeit.

³⁹⁹ Kalman (1982b, S. 162). Hervorhebung im Original.

⁴⁰⁰ Ebda., S. 171.

- (iv) Eine einzige Zahl bildet die Antwort eines deduktiven Prozesses, der von selbstevidenten Postulaten ausging.

Die Annahme von exakten Gesetzmäßigkeiten in probabilistischen Phänomenen analog zur Newtonschen Physik, die mit einer solchen Vorgehensweise angestrebt werde, habe sich jedoch längst als Illusion erwiesen. Abgesehen von "mathematischen Artefakten" wie dem Gesetz der großen Zahlen gäbe es keine universalen Gesetze für zufällige Phänomene – auch nicht in der Physik –, sondern stets nur solche, die von dem sie umfassenden System abhängen.⁴⁰¹ Solch eine Sichtweise hat tiefgreifende Konsequenzen:

"The implications of this situation for econometric strategy are devastating. Since the problem is to identify a system and since systems cannot be described in general by globally definable parameters, the whole idea of a parameter loses its (uncritically assumed) significance. [...] The *Jugendtraum* of econometrics, determining economically meaningful parameters from real data via dynamical equations supplied from economic theory, turns out to have been a delusion."⁴⁰²

Diese Kritik von Kalman hat in der Ökonometrie – soweit wir sehen – bislang keinerlei Rezeption gefunden.⁴⁰³ Auch wenn man ihr nicht in ihrer ganzen Konsequenz folgen will, so sollte die grundsätzliche Angemessenheit der Übertragung physikalischer auf ökonomische Entwicklungen dennoch hinterfragt werden. Auch in diesem Zusammenhang ist unseres Erachtens eine bayesianische Argumentation anstelle des von Kalman vorgeschlagenen systemtheoretischen Ansatzes legitim. Wir hatten am Ende von Abschnitt A 4 c darauf hingewiesen, daß in der statistischen Literatur der von Kalman entwickelte Algorithmus, der Kalmanfilter, zumeist in einem bayesianischen Rahmen interpretiert wurde. In diesem Zusammenhang werden keine Modelle mit festen Parametern postuliert, sondern anhand der Daten erfolgt eine permanente Revision unserer Einschätzung. Auch hiermit sind natürlich nicht alle Probleme gelöst: Zwar ist die Modellform erheblich flexibler, aber nach wie vor konstant. Mittlerweile haben jedoch auch in diesem Punkt (bayesianische) Überlegungen eingesetzt, die eine permanente Revision der Modellform ermöglichen.⁴⁰⁴

⁴⁰¹ Vgl. ebda., S. 172.

⁴⁰² Kalman (1982a, S. 26, 27). Die Berechnung eines konstanten Parameters, z. B. im Rahmen der Phillips-Kurve, bezeichnet er als "conceptual absurdity" (ebda.). Folgerichtig lehnt Kalman daher auch jede kausale Interpretation ab. Vgl. etwa Kalman (1982b, S. 177).

⁴⁰³ Sie spielt weder in den ökonometrisch orientierten Abhandlungen zu *state-space*-Modellen wie Aoki (1987), Harvey (1987) oder Harvey (1992) noch in den Ansätzen von Sims oder Hendry eine Rolle. Bei Leamer (1985, S. 255) findet sich eine kurze Erwähnung von Kalman (1982c); er bezeichnet Kalmans Ausführungen aber als eine "entirely different perspective."

⁴⁰⁴ Phillips (1995a), Phillips (1995b) und Phillips (1996).

“Die Uebernahme von Prinzipien der im strengen Sinn sog. ‘Wahrscheinlichkeitsrechnung’ kommt für die kausale Arbeit der Geschichte nicht nur, selbstverständlich, nicht in Betracht, sondern schon der Versuch einer analogen Verwertung ihrer Gesichtspunkte erheischt große Vorsicht.”

Max Weber (1906 [1968, S. 269])

Die Entwicklung der quantitativen Wirtschaftsgeschichte

In diesem Abschnitt wollen wir den Bogen schließen und nach der Behandlung methodologischer Konzepte der Statistik und Ökonometrie auf die Wirtschaftsgeschichte zurückkommen. Um die Bedeutung solcher Konzepte für die Wirtschaftsgeschichte und ihre Rezeption beurteilen zu können, müssen wir zunächst diese Disziplin näher eingrenzen. Daß es sich hierbei keineswegs um eine triviale Aufgabe handelt, zeigen die jüngst unter dem Titel “Wirtschafts- und Sozialgeschichte – Neue Wege? Zum wissenschaftlichen Standort des Faches” in der *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte* versammelten Diskussionsbeiträge, die einen Eindruck vom Selbstverständnis des Faches im deutschen Sprachraum vermitteln.¹ Eine vergleichbare Vielfalt zeigt sich bei einem Blick in die einschlägigen Lexikonartikel und Lehrbücher.² Insbesondere kann man nicht von der Entwicklung “der” Wirtschaftsgeschichte ausgehen, sondern muß nationale Besonderheiten berücksichtigen und noch stärker als im Bereich der Statistik und Ökonometrie das Wirken einzelner Personen in Rechnung stellen.³

Ein vollständiger wissenschaftshistorischer Erklärungsansatz für die Verwendung methodologischer Konzepte in der Wirtschaftsgeschichte müßte spätestens zu Beginn des 19. Jahrhunderts einsetzen und die Trennung der Disziplinen Statistik, Geschichte, Ökonomie und Soziologie analysieren. Auch eine solche – unseres Wissens bislang nicht vorliegende – Analyse könnte ein eigenes Buch füllen. Wir wollen uns hier auf einige wesentliche Aspekte beschränken, auch wenn wir dadurch Gefahr laufen, grob zu vereinfachen. Es muß allerdings darauf hingewiesen werden, daß es hier *nicht* unsere Intention ist, die Entwicklung der Wirtschaftsgeschichte oder ihrer Stellung in den Wirtschaftswissenschaften zu beschreiben

¹ Vgl. Köllmann et al. (1995).

² Man vergleiche nur die Artikel von Kellenbenz (1965), North (1968), Kocka (1972), Zorn (1982) oder Crafts (1991) sowie noch die unterschiedlichen Ansätze von Walter (1994) und Ambrosius et al. (1996).

³ “Alle Wirtschaftsgeschichtsschreibung in den westlichen Ländern präsentierte sich seit Ende des vorigen Jahrhunderts mit nationalen Unterschieden, erwachsen aus politischen, kulturellen und geistigen Traditionen. Gemeinsam war ihr gewöhnlich der institutionelle Rahmen, ihre Pflege in bestimmten, um bedeutende Köpfe gebildete Schulen, verbunden mit Lehrkanzeln, Instituten, Fachzeitschriften und Fachbuchserien und einer Schar von beflissenen Jungforschern.” Boelcke (1987, S. 7).

oder zu analysieren, sondern wir wollen lediglich die Anwendung quantitativer Methoden bzw. die Einstellung zu dieser Frage beleuchten.

Zur besseren Orientierung wollen wir dabei die deutsche, englische und amerikanische Entwicklung getrennt behandeln.

Die Etablierung der Wirtschaftsgeschichte als wissenschaftliche Disziplin ist eine Entwicklung unseres Jahrhunderts, obwohl das Interesse an historischen Wirtschaftsphänomenen zumindest bis in das 16. Jahrhundert zurückreicht. Zu den bekanntesten frühen Beispielen für ein solches Interesse zählen Untersuchungen von J. Bodin [1530-1596], W. Petty [1623-1687] oder im 18. Jahrhundert von S. de Vauban [1633-1707], P. de Boisguillebert [1646-1714] und R. Cantillon [1680-1734]. Wirtschaftsgeschichte wurde im 18. Jahrhundert in der Regel im Rahmen einer Staatsbeschreibung betrieben.⁴ In den klassischen Werken der englischen Nationalökonom A. Smith [1723-1790], T. Malthus [1766-1834] und D. Ricardo [1772-1823] finden sich ebenfalls wirtschaftsgeschichtliche Betrachtungen. Diese waren jedoch eher beispielhafter Natur; das Interesse war hier vorrangig auf die Gegenwartssituation gerichtet.⁵

Die Entwicklung in Deutschland

Im 19. Jahrhundert kann man einen ersten Vorläufer der Wirtschaftsgeschichte in Deutschland in den Arbeiten von Wilhelm Roscher [1817-1894], Carl Knies [1821-1898] und Bruno Hildebrand [1812-1886] sehen, die später von A. Wagner als "Ältere Historische Schule" der Nationalökonomie bezeichnet wurden.⁶ Roscher war der Ansicht, daß einer ökonomischen Analyse historische Methoden zugrundegelegt werden sollten. Eine derartige Analyse könne nur im Verbund mit der Rechts-, Staats- und Kulturgeschichte sinnvoll sein. Was er genau unter einer historischen Methode verstand, bleibt allerdings weitgehend unklar; in der Regel beschränkte sich Roscher auf die beispielhafte Verwendung historischer Analogien.⁷ B. Hildebrand, Begründer der *Jahrbücher für Nationalökonomie und Statistik*, legte dagegen größeren Wert auf eine Ergänzung der abstrakten, seiner Ansicht nach spekulativen Methode der klassischen ökonomischen Theorie durch die

⁴ Siehe zu dieser Entwicklung Gras (1927, S. 13-18) sowie mit Ausführungen über die Historiographie der Wirtschaftsgeschichte seit der Antike Beutin/Kellenbenz (1973, S. 123-148).

⁵ Dies betont zu Recht Kellenbenz (1965, S. 134).

⁶ Für ein tieferes Verständnis müßte hier eigentlich die Entwicklung der Statistik in Deutschland im 19. Jahrhundert behandelt werden, da die historische Schule viel stärker auf diese Entwicklung rekurrierte als auf die englische Tradition. Diese "Universitätsstatistik" lehnte überwiegend das Postulieren von Gesetzmäßigkeiten im Sinne Quetelets oder auch nur der Relevanz von Regularitäten ab. Dennoch fielen auch hier die Ideen Quetelets zunächst auf fruchtbaren Boden. Ernst Engel, der Gründer des Berliner Statistischen Seminars, sah sich z. B. in dessen Tradition. G. Rümelin, einer der bedeutendsten Vertreter der deutschen Statistik im 19. Jahrhundert, stand Quetelet nach anfänglicher Zustimmung letztendlich ablehnend gegenüber. Vgl. dazu allgemein Porter (1987), Hacking (1987) sowie Porter (1986, S. 162-171, 179).

⁷ Wir stützen uns hier vor allem auf Weber (1903-1906).

Statistik. Eine systematische Anwendung ist er jedoch schuldig geblieben. K. Knies führte die Gedankengänge von Roscher und Hildebrand im wesentlichen fort.

Der erste wichtige Anstoß für die Wirtschaftsgeschichte wurde von der sogenannten "Jüngeren Historischen Schule der Nationalökonomie" gegeben. Die ältere historische Schule wies auf die Notwendigkeit der Berücksichtigung historischer Umstände bei der Konstruktion ökonomischer Theorien hin. Sie lehnte die Verwendung von Theorien lediglich in einer rein deduktiven Form – wie etwa in der Tradition Ricardos – ab. Die jüngere historische Schule, die von G. Schmoller begründet wurde,⁸ ging darüber hinaus: Ihrer (Schmollers) Ansicht nach hatten abstrakte ökonomische Theorien in der Wirtschaftswissenschaft keinen Platz; diese sollte sich vielmehr aus der Generalisierung historischer Einzeluntersuchungen konstituieren.⁹

Eine solche Einstellung stand im Widerspruch zur zeitgenössischen Ökonomie. Seit John St. Mills *System der Logik* wurde eine deduktive, von abstrakten Theorien ausgehende Vorgehensweise allgemein höher eingeschätzt als eine induktive, Beobachtungen verallgemeinernde Forschung. Durch die Arbeiten von Menger, Walras, Jevons, Cournot und anderen war bereits ein Abstraktionsniveau erreicht, das von universalen und allgemeingültigen ökonomischen Gesetzen ausging. Mengers 1883 erschienene, umfangreiche "Untersuchungen über die Methode der Sozialwissenschaften und der Politischen Ökonomie insbesondere" liefen darauf hinaus, selbst für die Falsifizierung von Theorien lediglich logische Kriterien gelten zu lassen. Ein solcher Ansatz mußte natürlich von Schmoller als Bedrohung empfunden werden. Wir wollen diese Kritik und die Reaktion Mengers hier nicht im einzelnen wiedergeben.¹⁰ Von Interesse ist für unseren Zusammenhang die Tatsache, daß Schmoller zwar eine induktive Vorgehensweise forderte, darunter aber keine statistischen, sondern qualitative Forschungen verstand. Statistische Gesetze im Sinne Quetelets und seiner Nachfolger lehnte er ebenso ab wie die universalen ökonomischen Gesetze Mengers und betonte auch hier die Zeit- und Ortsgebundenheit von Massenphänomenen in der Gesellschaft. Noch in seinem

⁸ Gustav von (seit 1908) Schmoller [1838-1917] promovierte nach einem Studium der Kameralwissenschaften in Tübingen mit einer "Untersuchung der volkswirtschaftlichen Anschauungen zur Reformationszeit" und arbeitete anschließend unter Gustav Rümelin beim Württembergischen Statistischen Amt an einer Auswertung der württembergischen Gewerbebeziehung von 1861, dessen Veröffentlichung ihn von der Notwendigkeit empirischer Arbeit in den Wirtschaftswissenschaften überzeugte. Nach einer außerordentlichen Professur an der Universität Halle 1864 und maßgeblicher Mitbeteiligung an der Gründung des *Vereins für Socialpolitik* wurde er 1872 an die Reichsuniversität Straßburg berufen, 1882 nach Berlin. Seit 1881 war er Herausgeber der *Jahrbücher für Gesetzgebung, Verwaltung und Volkswirtschaft im Deutschen Reich* ("Schmollers Jahrbuch"), das die deutsche Nationalökonomie für Jahrzehnte maßgeblich beeinflusste. Seit 1887 begann er die Herausgabe der *Acta Borussica*, einer Edition von Quellen zur preussischen Staats- und Wirtschaftsverwaltung.

⁹ Vgl. Nardinelli/Meiners (1988). Ihr gehörten daneben Eberhard Gothein, Georg Friedrich Knapp und Georg Schanz an. Vgl. Boelcke (1987, S. 5).

¹⁰ Siehe dazu insbesondere die Artikel des Sonderheftes des *Journal of Institutional and Theoretical Economics* (*Views and comments on Gustav Schmoller and the Methodenstreit*) 144 (1988).

Beitrag „Volkswirtschaft, Volkswirtschaftslehre und -methode“ für die dritte Auflage des *Handwörterbuch der Staatswissenschaften* schrieb er: „Wir verstehen unter statistischen Gesetzen die zahlenmäßige durch statistische Massenbeobachtung konstatierten Regelmäßigkeiten von Geburten, Eheschließungen, Todesfällen, Verbrechen und ähnliche soziale Einrichtungen.“¹¹ Die Kenntnis solcher statistischen Gesetze sei jedoch nicht überzubewerten, wie dies etwa durch Quetelet oder Buckle geschehen sei: „Man vergaß, auf wie wenige Fragen sich diese Massenzählung anwenden läßt, wie roh und äußerlich die Fragestellung bleibt, wie die Statistik sich überwiegend im Gebiete des Naturgeschehens bewegt.“¹² Er wollte daher die Anwendung des Gesetzesbegriffes etwa – auf Simmel verweisend – für den Anteil der Selbstmörder einer Gesellschaft gelten lassen, nicht jedoch für ökonomische Erscheinungen: „Ob man auch von einem Gesetz des abnehmenden Bodenertrags oder von einem Malthusschen Bevölkerungsgesetz sprechen könne, möchte ich dagegen meinerseits sehr bezweifeln.“¹³ Die bereits über ein Jahrzehnt zuvor erschienenen Arbeiten von Edgeworth oder Yule¹⁴ wurden mit keinem Wort erwähnt.

Es wird gewöhnlich darauf verwiesen, daß Schmoller und die jüngere historische Schule auch in methodischer Hinsicht das Instrumentarium der allgemeinen Geschichtswissenschaft übernahmen. Insofern ist es erstaunlich, daß der Beginn der deutschen Wirtschaftsgeschichte als Disziplin, die aus der allgemeinen Geschichtswissenschaft hervorging, sich gegen die historische Schule der Nationalökonomie richtete. Um diese Entwicklung zu verstehen, erweisen sich zunächst einige Bemerkungen zur allgemeinen Geschichtswissenschaft als notwendig.

Die deutsche Historie hatte sich aus einer gelehrten und einer literarischen Wurzel heraus entwickelt, die in einem längeren Prozeß miteinander verschmolzen.¹⁵ Herausragendster Vertreter dieser Synthese war zweifellos Leopold von Ranke [1795-1886], der „Vater der modernen Geschichtswissenschaft“. Demnach war die Geschichtsschreibung eine Wissenschaft mit eigenen Methoden (Quellenkritik), aber darüber hinaus ebenso literarischen Kriterien unterworfen. Die großen Historiker des 19. Jahrhunderts (Ranke, Burckhardt, Gervinus, Droysen, Treitschke, Mommsen) bemühten sich daher, auch für ein breites Publikum zu schreiben.¹⁶ Nach allgemeiner Grundüberzeugung bestand die Aufgabe der Geschichtsschreibung in einer strengen Darstellung der Tatsachen, „wie es wirklich gewesen ist“. Sie war einer spezifischen Definition des Historismus verpflichtet, die später vor allem von W. Dilthey, W. Windelband und H. Rickert wissenschaftstheoretisch begründet wurde. Demnach war der Historismus

¹¹ Schmoller (1911, S. 485).

¹² Ebda. Auch Knapp (1869) stellte sich explizit gegen Quetelet.

¹³ Schmoller (1911, S. 486). Heaton (1949) beschreibt ihn als „ever critical of anyone who believed he had found in some statistics a new bit of truth.“

¹⁴ Siehe oben, Abschnitt A 3 c.

¹⁵ Vgl. zum folgenden Iggers (1996).

¹⁶ Sichtbarster Ausdruck ist sicher die Verleihung des Literaturnobelpreises an Mommsen 1902.

“[...] eine Weltanschauung und eine Wissenschaftskonzeption, die im Gegensatz zum Hegelschen oder Marxschen Glauben an Gesetzmäßigkeiten in der Geschichte die spontanen, unberechenbaren Elemente der menschlichen Freiheit und Kreativität betont. Diese benötigen eine Logik der Forschung, des Verstehens menschlicher Zusammenhänge, die grundsätzlich anders ist als die der Naturwissenschaften.”¹⁷

Es mag vor diesem Hintergrund nicht überraschen, daß statistische Ansätze *prinzipiell* auf Ablehnung stießen. In einer programmatischen Besprechung von Buckles *Geschichte der Civilisation*¹⁸ durch J. G. Droysen wurde die Geschichte gerade erst dadurch in den Rang einer Wissenschaft erhoben, daß sie Erscheinungen in den Mittelpunkt des Interesses stelle, für die eine physikalische oder mathematische Betrachtungsweise nicht geeignet ist. Hier gehe es nicht um Deduktion oder Induktion, sondern um ein “Verstehen”.¹⁹ Diese richtungsweisende Grundeinstellung bringt ein von Droysen angeführtes Beispiel besonders deutlich zum Ausdruck: Wenn man eine menschliche Existenz als Summe zweier Bestandteile, $A = a + x$, auffasse, in der a den durch äußere Umstände bedingten Teil seiner Existenz und x “sein eigenes Zuthun, das Werk seines freien Willens” bezeichne, dann möge dieses x zwar verschwindend klein sein, es markiere jedoch den entscheidenden Teil:

“Mag immerhin die Statistik zeigen, daß in dem bestimmten Lande so und so viele uneheliche Geburten vorkommen, mag in jener Formel $A = a + x$ dieß a alle die Momente enthalten, die es ‘erklären’, daß unter tausend Mädchen 20, 30, wie viele es denn sind, unverheiratet gebären, – jeder einzelne Fall der Art hat seine Geschichte und wie oft eine rührende und erschütternde, und von diesen 20, 30 Gefallenen wird schwerlich auch nur eine sich damit beruhigen, daß das statistische Gesetz ihren Fall ‘erkläre’; in den Gewissensqualen durchweinter Nächte wird sich manche von ihnen sehr gründlich überzeugen, daß in der Formel $A = a + x$ das verschwindend kleine x von unermeßlicher Wucht ist, daß es den ganzen sittlichen Werth des Menschen, das heißt seinen ganzen und einzigen Werth umschließt.”²⁰

Für die Geschichtswissenschaft bedeutete diese Haltung für Jahrzehnte eine an politischen Ereignissen und Personen orientierte, mit quellenkritisch philologischen Methoden arbeitende, narrative Ausrichtung. Gegen Ende des 19. Jahrhunderts wiesen erste Anzeichen jedoch auf eine “Krise” hin.²¹ Es mehrten sich Stimmen, die eine stärkere Berücksichtigung von Aspekten der “Kultur” und “Gesellschaft” forderten. Über zehn Jahre nach der – von der Fachhistorie nicht

¹⁷ Iggers (1996, S. 19).

¹⁸ Das Werk war bereits 1864 in zweiter deutscher Auflage erschienen und erlebte bis 1901 fünf weitere Auflagen.

¹⁹ Vgl. Droysen (1863, S. 7, 12f).

²⁰ Ebda., S. 14.

²¹ Siehe dazu etwa den kurzen Überblick von Iggers (1996). Nach Schleier (1988b, S. 16f) war die Diskussion über die Definition und Ausrichtung der Geschichtswissenschaft nicht auf Deutschland beschränkt: “Immer häufiger wurde das theoretisch-methodologische Instrumentarium der Geschichtswissenschaft als unzureichend empfunden. In zahlreichen Ländern begann man etwa zur gleichen Zeit Ausschau nach neuen Konzepten, Geschichtstheorien und Methoden insbesondere in Richtung auf eine Sozialgeschichte zu halten.”

zur Kenntnis genommenen – Kontroverse zwischen Schmoller und Menger kam es auch in der Geschichtswissenschaft zu einem "Methodenstreit", der mit dem Namen K. Lamprechts verbunden ist.²²

Dieser Methodenstreit begann durch einen Verriß von Lamprechts *Deutscher Geschichte* in der *Historischen Zeitschrift* durch von Below und zog sich bis zur Jahrhundertwende mit Angriffen von F. Rachfahl, M. Lehman und F. Meinecke hin.²³ Auch hier wollen wir dem Streit nicht im einzelnen nachgehen, sondern lediglich auf die für unseren Zusammenhang relevanten Aspekte hinweisen.

Lamprecht wies wie schon T. Buckle auf die Bedeutung materieller Faktoren für die allgemeine historische Entwicklung hin. Ihm waren die Schwächen von Buckles Werk durchaus bewußt, diese führten aber nicht zu einer prinzipiellen Ablehnung, wie etwa bei Droysen 1863. Für ihn bestand die wissenschaftliche Geschichtsschreibung vor allem darin, empirische Regularitäten in der progressiven Entwicklung von Kulturen festzustellen. Er rezipierte zwar Quetelet²⁴ und W. Lexis, seine Anwendung statistischer Überlegungen auf "Kulturzeitalter" hätte aber wohl kaum deren Zustimmung gefunden:

"Darnach [sic!] kommt es jetzt darauf an, die diesen Zeitaltern zugrunde liegenden Kausalverhältnisse, Notwendigkeiten, empirischen Gesetze tatsächlich nachzuweisen. Es ist eine Aufgabe, zu deren Lösung die analogen Forschungen auf dem Gebiete der Biologie schon längst einigen Anhalt gewähren können. Dort ist zunächst geschehen, was durch die Aufstellung der Kulturzeitalter nun auch für die Geschichte gegeben ist: Die bisher ganz individuellen Entwicklungsregeln für die einzelnen organischen Spezies sind auf allgemeine Entwicklungsregeln der Arten reduziert worden."²⁵

²² Karl Lamprecht [1856-1915] studierte Geschichte in Göttingen, München und Leipzig, promovierte 1878 mit einer Arbeit über "Frankreichs wirtschaftliche Verhältnisse im 11. Jahrhundert" und widmete sich zunächst der rheinischen Kultur- und Wirtschaftsgeschichte. Nach Erscheinen seines ersten Hauptwerkes "Deutsches Wirtschaftsleben im Mittelalter" 1885/86 wurde er 1885 Extraordinarius in Bonn; 1890 folgte er einem Ruf nach Marburg, 1891 nach Leipzig, wo er Vorlesungen vor bis zu 700 Studenten hielt und insgesamt 140 Dissertationen vergab. Zwischen 1890 und 1895 erschien sein zweites Hauptwerk, die fünfbandige "Deutsche Geschichte". Eine Reihe seiner theoretischen Schriften sind in Schleier (1988a) wiederabgedruckt.

²³ Von Below (1893). Georg von Below [1858-1927], u. a. Schüler von Droysen und Treitschke, promovierte über die "Entstehung des ausschließlichen Wahlrechts der Domkapitel" und war Professor für mittlere und neuere Geschichte, zuletzt in Münster. Von Belows Kritik bestand aus einer Bemängelung zahlreicher Detailfehler. Lamprecht antwortete seinen Kritikern in Lamprecht (1896). Vgl. Schleier (1988b, S. 20f, die Angaben zu den Kritiken S. 143).

²⁴ So sieht er etwa Quetelets Verdienst darin, daß er das von J. und D. Bernoulli, Laplace, Poisson und Lacroix auf dem Wege mathematischer Analysis bewiesene Gesetz der großen Zahlen "praktisch endlich [...] durch Experimente bestätigt" habe. Lamprecht (1896-1897 [1988, S. 260]).

²⁵ Ebda., S. 263. Er zitierte hierbei zwar Mayr (1877), bevorzugte aber in Anlehnung an Wilhelm Wundts *Logik* den Ausdruck "empirisches Gesetz". Vgl. zu Lamprechts Kausalitätsbegriff Wise (1987, S. 404-406). Wise verallgemeinert allerdings die in Lamprecht (1896-1897) geäußerten Thesen in unzulässiger Weise.

Die Antwort auf die genaue Durchführung eines solchen Nachweises blieb Lamprecht allerdings schuldig; diese hätte wohl auch nur abenteuerlich ausfallen können. Es läßt sich wohl festhalten, daß Lamprechts Ausführungen, von Max Weber als 'spekulativer Unsinn' bezeichnet, nicht gerade dazu beitrugen, die Anwendung statistischer Konzepte in der Geschichtswissenschaft zu fördern.

Sowohl gegen diese als auch gegen die von Schmoller und seinen Schülern vertretende Auffassung distanzierte sich die etwa seit der Jahrhundertwende im Entstehen begriffene eigenständige Wirtschaftsgeschichte.²⁶ Das Erkenntnisziel der Wirtschaftsgeschichte sollte nicht in der Verallgemeinerung ökonomischer Erscheinungen zu theoretischen Kategorien oder Konstrukten bestehen und schon gar nicht durch statistische Methoden erreicht werden. Die 1903 von Historikern gegründete *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte* machte bereits im Vorwort der ersten Ausgabe deutlich, daß man sich von der "Behandlung der Probleme der theoretischen Nationalökonomie wie von Fragen der Social- und Volkswirtschaftspolitik der Gegenwart fernhalten" wollte.²⁷

G. von Below übte seit 1915 als maßgeblicher Herausgeber der *Vierteljahrschrift* maßgeblichen Einfluß auf die inhaltliche Ausrichtung der Disziplin aus. Die institutionelle Etablierung der deutschsprachigen Wirtschaftsgeschichte setzte im zweiten Jahrzehnt unseres Jahrhunderts ein. Die ersten wirtschaftsgeschichtlichen Lehrstühle in Deutschland wurden 1919 an der Handelshochschule/Universität Köln (Bruno Kuske) und 1920 in München eingerichtet.²⁸ In München wurde 1921 unter dem Einfluß Max Webers das Seminar für Wirtschaftsgeschichte gegründet. Erster Leiter wurde Jakob Strieder.²⁹ Sowohl Kuske als auch Strieder waren Historiker und keine Ökonomen.

Alfons Dopsch [1868-1953], wie von Below Mediävist, lehnte 1920 einen Ruf als Nachfolger Schmollers ab und gründete 1922 das Seminar für Wirtschafts- und Kulturgeschichte an der Wiener Universität. Dopsch begriff "als Hauptzweck seiner Forschung die Falsifizierung der ökonomisch-historischen Theorien [...]."³⁰ Und als E. Salin 1921 die Methoden und Aufgaben der Wirtschaftsgeschichte definierte, sah er diese gerade im Gegensatz zu den systematischen Sozialwissenschaften:

²⁶ Nach Mooser (1990, S. 91f) führte weniger ein "fundamentaler, weltanschaulich-politischer Gegensatz, sondern professionspolitisches Selbstverständnis im Verein mit disziplinärer Spezialisierung methodischer Standards [...] zur Trennung der Historiker und (deutschen) Nationalökonomien". Ein solches Argument unterschätzt jedoch den evolutischen Charakter von Schmollers Ansatz. Iggers (1996, S. 31) kritisiert zu Recht, daß die Betonung der Quellenkritik bei Schmoller "[...] ergänzt [wurde] um eine evolutionäre Stufenlehre, deren spekulativer Charakter in einem gewissen Widerspruch zu seiner sorgfältigen Quellenarbeit stand. Generell kann man sagen, daß die Arbeiten Schmollers und der Jünger der Historischen Schule der Nationalökonomie auf theoretischen und methodischen Voraussetzungen beruhten, die sie nie kritisch oder systematisch expliziert hat."

²⁷ Zitiert nach Mooser (1990, S. 88).

²⁸ Vgl. Borchardt (1987, S. 17).

²⁹ Vgl. Zorn (1974, S. 15).

³⁰ Mooser (1990, S. 91). Siehe in methodologischer Hinsicht etwa Dopsch (1927).

„Wohl aber soll gesagt sein, daß der Historiker, gerade wenn ihm der Theoretiker und der Soziologe feste Begriffe bieten, die Aufgabe hat, seinen Blick auf das Einmalige und Eigene zu richten, Analogien nicht vor auszusetzen, Zeitliches nicht als ewig, Einmaliges nicht als immer vorhandenes zu betrachten.“³¹

Auch in den Arbeiten der Ökonomen oder Soziologen, die sich wirtschaftshistorischen Fragestellungen zuwandten (M. Weber, W. Sombart) spielten quantitative Methoden keine Rolle. Dafür erlebte durch keinen geringeren als Schumpeter der Ansatz von Schmoller eine Wiederaufnahme. 1926 verfaßte Schumpeter seinen berühmten Aufsatz *Gustav von Schmoller und die Probleme von heute*,³² in dem er zunächst eine Rehabilitation Schmollers forderte und die sich durch den „Methodenstreit“ verfestigten Fronten versöhnen wollte.³³ Anlaß war eine seiner Meinung nach bestehende Gefahr, daß sich die verhängnisvolle Entwicklung des Methodenstreits nach dreißig Jahren in den USA wiederholen könnte.³⁴ Nach Ansicht Schumpeters habe man Schmollers Einstellung zu Theorien lange Zeit, insbesondere nach dem Methodenstreit, mißverstanden.³⁵ Er habe sich lediglich gegen eine abstrakte Theorievorstellung, gegen den Glauben an die Möglichkeit einer nahezu vollständigen Formalisierung und Mathematisierung der Wirtschaftswissenschaft gewandt, andererseits habe sein Werk aber durchaus – in einem weiteren Sinne – theoretische Elemente enthalten.³⁶ Auch die ihm insbesondere von Max Weber vorgeworfene mangelnde Wertfreiheit³⁷ sei in den meisten Schriften Schmollers nicht zu beobachten gewesen. Schumpeter ging über eine reine Rehabilitation Schmollers hinaus: Er verband mit dessen Ansatz sogar eine Forderung für ein Programm der Zukunft.³⁸ Schmollers Berücksichtigung dessen, was die Ökonomen mittlerweile zu vernachlässigen pflegten, die konkreten Umstände,

³¹ Salin (1921, S. 500). Der erste Teil dieses Beitrags bestand aus einer kritischen Ablehnung von F. Lists Lehre der Wirtschaftsstufen. Der zweite Teil, „Von den Aufgaben der Wirtschaftsgeschichte“, war Salins Habilitationsvortrag an der Universität Heidelberg. Edgar Salin [1892-1974] wurde nach Promotion und Habilitation in Heidelberg 1927 Professor der Staatswissenschaften in Basel.

³² Schumpeter (1926).

³³ Siehe dazu auch Swedberg (1994, S. 120-126).

³⁴ Er wies in diesem Zusammenhang auf die von Mitchell gegenüber der *American Economic Association* geäußerte Kritik hin. Siehe oben, Abschnitt B 3 b.

³⁵ Auch er selbst habe in seinem 1914 erschienenen Buch *Epochen der Dogmen- und Methodengeschichte* diesen Fehler begangen. Vgl. Schumpeter (1926, S. 355).

³⁶ In diesem Zusammenhang ist auch Spiethoffs Forderung nach einer „anschaulichen Theorie“ zu erwähnen. Spiethoff war Assistent bei Schmoller und später Herausgeber von *Schmollers Jahrbuch*.

³⁷ Dieser dritte Methodenstreit, der unter der Bezeichnung „Werturteilsstreit“ in die Wissenschaftsgeschichte eingegangen ist, hatte den Gegensatz zwischen positiver und normativer Wissenschaft zum Thema und war vor allem gegen Schmoller gerichtet. Er ging zwar über eine ontologische Kontroverse hinaus, da sich in diesem Zusammenhang auch die Frage stellte, inwiefern die Verwendung bestimmter Methoden den Charakter einer Wissenschaft bestimmt und nicht umgekehrt. Da jedoch die konkrete Verwendung von Methoden nicht zur Debatte stand, wollen wir auf dieses Problem hier nicht weiter eingehen.

³⁸ Vgl. Schumpeter (1926, S. 360).

die sozialen Strukturen, generell die institutionelle Seite, statistische Daten, kurzum die Bereiche Geschichte und Statistik, müsse die Ökonomie in Zukunft wieder stärker in den Mittelpunkt rücken. Schmoller selbst habe diese Vorgehensweise nicht nur als Programm gefordert, sondern auch, wie Schumpeter an mehreren Beispielen aufzählte, erfolgreich realisiert.³⁹ Diese Verbindung historischer, theoretischer und statistischer Elemente, die er unter der Bezeichnung *Socialökonomik* subsumierte, sollte Schumpeter selbst über ein Jahrzehnt später seiner monumentalen Konjunkturzyklenuntersuchung zugrundelegen.⁴⁰ Wie wir gesehen hatten, ging er in bezug auf die Forderung nach quantitativen Untersuchungen zwar über Schmollers eigene Vorstellungen hinaus, blieb in der Umsetzung jedoch zurückhaltend.⁴¹

Im Rahmen der Wirtschaftsgeschichte setzte sich dagegen die atheoretische Detailforschung bis in die Nachkriegszeit fort.⁴² Die bedeutendsten Vertreter des Faches waren einer Tradition verpflichtet, die sich eher an der Geschichtswissenschaft als an der zunehmend mathematischen Nationalökonomie orientierte. Dies gilt insbesondere für F. Lütge und H. Kellenbenz, die das Fach Wirtschaftsgeschichte in Deutschland nach dem Krieg maßgeblich bestimmten.⁴³ Kellenbenz, wohl auch international der einflußreichste deutsche Wirtschaftshistoriker,⁴⁴ stellte

³⁹ Wie Swedberg (1994, S. 123) hervorhebt, ist die Forderung Schumpeters in Anlehnung an Schmoller eine fundamentale: Es ging nicht um eine *Ergänzung* der zeitgenössischen Ökonomie um historische Elemente, sondern um eine *Umwälzung* der Wirtschaftswissenschaft in eine solche, die alle Phänomene aus ihrer historischen Entstehung heraus in umfassender Weise erklärt.

⁴⁰ Siehe oben, Abschnitt B 2 d.

⁴¹ Swedberg (1994, S. 379) ist der Ansicht: "Schumpeter hätte der gegenwärtigen Tendenz zur Anwendung verfeinerter quantitativer Methoden bei der wirtschaftshistorischen Forschung höchstwahrscheinlich sehr kritisch gegenübergestanden. Schließlich war er durch die andersartige Tradition der deutschen Historiographie geprägt worden."

⁴² Die Wirtschaftsgeschichtsschreibung der DDR, auf die wir hier nicht näher eingehen wollen, hatte zwar eine andere Themenausrichtung und eine weitaus größere politische Bedeutung; ihre methodologischen Grundlagen wiesen jedoch keine bedeutenden Unterschiede zur westdeutschen Entwicklung auf. Vgl. zu dieser Entwicklung Grabas (1991).

⁴³ Friedrich Lütge [1901-1968] begann ein Studium der Volkswirtschaftslehre, entwickelte aber bereits früh ein ausgeprägtes historisches Interesse und promovierte über "Die Bauernbefreiung in Wernigerode". Starken Einfluß übte G. von Below aus. Lütge hatte in München zunächst (seit 1947) einen Lehrstuhl für Wirtschaftsgeschichte und Volkswirtschaftslehre, ab 1960 einen Lehrstuhl für Wirtschafts- und Sozialgeschichte. Bis 1968 war er Vorsitzender der 1961 gegründeten *Gesellschaft für Sozial- und Wirtschaftsgeschichte*. Borchardt (1970, S. 3) formuliert Lütges Einstellung zur Entwicklung der Wirtschaftswissenschaft wie folgt: "Die zunehmende Berücksichtigung der Mathematik zur Formulierung theoretischer Gedankengänge in der Volkswirtschaftslehre kritisierte er wohl nicht nur aus Einsicht in die Problematik. Er hatte einfach einen schlechten Schulunterricht in Mathematik genossen und die Faszination der Formel, das wirkliche Vergnügen an der formalen Schönheit von Modellen nie verspüren dürfen."

⁴⁴ Hermann Kellenbenz [1913-1990] studierte Geschichte und Kunstgeschichte, promovierte 1938 über die Domäne Holstein-Gottorf im 17. Jahrhundert, habilitierte sich kurz vor Kriegsende über "Sephardim an der unteren Elbe" und wurde nach mehreren Stationen 1970 Professor für "Geschichte mit besonderer Betonung der Wirtschafts- und Sozialgeschich-

noch 1965 die Forderung, "[...] *Wirtschaftsgeschichte als Geistesgeschichte* zu betreiben".⁴⁵ Eine Ausnahme bildete hier lediglich W. Abel, der 1935 mit einer bis heute als Standardwerk geltenden Arbeit über *Agrarkrisen und Agrarkonjunktur* Ansätze einer quantitativen Wirtschaftsgeschichte entwickelte.⁴⁶ Abel beschrieb seine Intention wie folgt:

"Was bedeuten die [...] langfristigen Schwankungen der Getreidepreise? Sind sie lediglich Folgen schwankenden Geldvorrates der mitteleuropäischen Wirtschaft, oder sind sie vielleicht auch Symptom einer wechselnden Spannung zwischen der Bevölkerung und dem Nahrungsspielraum? [...] So sollen einmal die vielberufenen 'Entwicklungsgesetze' eines Malthus und Ricardo an den Tatsachen der Geschichte überprüft werden, zum anderen, und das ist das Wesentlichere, historische Entwicklungsreihen mit Hilfe eines ihnen adäquaten Begriffs- und Beziehungsschemas auf ihren Sinngehalt gedeutet werden."⁴⁷

Methodisch beschränkte sich Abel hier wie in den folgenden Arbeiten auf graphische Darstellungen von Preisreihen, die allenfalls anhand gleitender Mittelwerte zusammengefaßt wurden. Von dem zwei Jahrzehnte zuvor erschienenen, hypothesengenerierenden Ansatz eines Mitchell oder gar dem Vorgehen von Moore⁴⁸ war auch diese Arbeit weit entfernt. Seine Schüler blieben in methodischer Hinsicht eher hinter seinem Ansatz zurück.⁴⁹

Die Entwicklung in England

In England verfestigte sich die Trennung der Disziplinen Wirtschaftstheorie und Wirtschaftsgeschichte zwischen den 1870er und 1920er Jahren. Aus verschiedenen Gründen schufen die englischen Wissenschaftler jedoch keine Schulen wie die

te" in Erlangen-Nürnberg. Nach dem Tode F. Lütges wurde er Vorsitzender der deutschen *Gesellschaft für Sozial- und Wirtschaftsgeschichte*, kurze Zeit später Vorstandsmitglied der *International Association of Economic History*. Sein Schüler W. Zorn (1991, S. 4) charakterisiert ihn treffend: "Hermann Kellenbenz war und blieb durch und durch ein quellennah exakter Historiker, im positiven Sinne ein 'Archivmensch', der hinter der Wirtschaftsgeschichte die Geistesgeschichte des Unternehmertums sah, während er der modernen ökonomischen Theorie, ihren Modellen und rekonstruierten statt unmittelbar überlieferten Zahlen mit großen Vorbehalten gegenüberstand."

⁴⁵ Kellenbenz (1965, S. 125). Hervorhebung im Original.

⁴⁶ Wilhelm Abel [1904-1985] promovierte nach einem Studium der Rechts- und Staatswissenschaften in Marburg, München und Kiel 1929 in Kiel mit einer Arbeit über "Die Träger des deutschen Getreidehandels", die aus seiner Mitarbeit bei der Edition der *Acta Borussica* hervorging. Nach Assistententätigkeiten habilitierte er sich 1935 an der Universität Frankfurt mit der mittlerweile klassischen Arbeit "Agrarkrisen und Agrarkonjunktur in Mitteleuropa vom 13. - 19. Jahrhundert" und erhielt dort die *venia legendi* für Volkswirtschaftslehre. Nach Tätigkeiten in Frankfurt/M. und Königsberg war er seit 1946 an der Universität Göttingen tätig, zunächst am Institut für Agrarpolitik, zuletzt bis 1973 als Direktor des Instituts für Wirtschafts- und Sozialgeschichte.

⁴⁷ Abel (1935, S. 11). Hervorhebung von uns.

⁴⁸ Siehe oben, Abschnitt B 2 b.

⁴⁹ Lediglich W. Achilles (1959) ging über Abel hinaus und berechnete Korrelationskoeffizienten zwischen Preisreihen. In methodischer Hinsicht ging diese Arbeit jedoch hinter Hooker (1901) zurück.

deutsche historische Schule.⁵⁰ Ebenso wie in Deutschland entwickelte sich eine wirtschaftshistorische Betrachtungsweise aus einer Kritik an der deduktiv ausgerichteten, hier neoklassischen Wirtschaftstheorie.

Für unseren Zusammenhang sind insbesondere die Arbeiten von T. Tooke und W. Newmarch (1838-1857) sowie J. E. T. Rogers (1866-1902) von Bedeutung. Tooke und Newmarch untersuchten Preise von 1793 bis 1857 und gelangten zu dem Ergebnis, daß die Höhe der Preise nicht vorrangig von der im Umlauf befindlichen Geldmenge, sondern zum überwiegenden Teil durch nicht endogen ökonomische Faktoren wie Kriege oder die Witterung beeinflusst wurden.⁵¹ In methodischer Hinsicht unterschied sich diese Arbeit nicht von anderen zu dieser Zeit. Sie war im Gegensatz zu den bisherigen empirisch-statistischen Untersuchungen jedoch von dem Versuch einer systematischen Berücksichtigung der historischen Umstände gekennzeichnet. Sie wurde durch das noch gewaltigere und durch sie inspirierte, zwischen 1866 und 1902 erschienene siebenbändige Werk *History of Agriculture and Prices in England* von J. E. T. Rogers weitergeführt.⁵²

Rogers hatte seine wissenschaftstheoretischen Ansichten in einem *Manual of Political Economy* dargelegt, in dem er auf die Bedeutung historischer Untersuchungen für das Verständnis gegenwärtiger ökonomischer Probleme hinwies; sein Name wird heute jedoch hauptsächlich mit der *History* verbunden. Rogers Arbeiten wiesen zwar auf die Bedeutung statistischer Untersuchungen für die Wirtschaftsgeschichte hin,⁵³ langfristig gesehen hatte sein Ansatz jedoch ebenso geringen Einfluß auf die englische Wirtschaftshistoriographie wie die Arbeiten von Tooke und Newmarch, die statistischen Analysen von W. S. Jevons oder die ersten zeitreihenanalytischen Versuche.⁵⁴

Die Entwicklung nahm hier einen ähnlichen Verlauf wie in Deutschland. Statistischen Angaben wurde allenfalls illustrativer Charakter zugebilligt. Während sich in Deutschland nach dem ersten Methodenstreit 1883/84 die Fronten zwischen theoretischer und historischer Nationalökonomie bereits verhärtet hatten, folgte hier eine solche Positionsabgrenzung ein knappes Jahrzehnt später, als W. Cunningham der ahistorischen Betrachtungsweise A. Marshalls eine "perversion of economic history" vorwarf.⁵⁵ Damit waren die Grenzen abgesteckt, und kurze Zeit

⁵⁰ Siehe dazu allgemein Koot (1987).

⁵¹ Vgl. ebda., S. 64. Die Bände erschienen 1858/59 in deutscher Übersetzung.

⁵² James E. Thorold Rogers [1823-1889] studierte in Oxford, wurde zum anglikanischen Priester geweiht, entschloß sich dann jedoch zu einer weltlichen Karriere und wandte sich ökonomischen Problemen zu. 1859 wurde er *Tooke Professor of Statistics and Economic Science* am *King's College* in London, drei Jahre später erhielt er einen Lehrstuhl für politische Ökonomie in Oxford, mußte diesen aber 1868 wieder abgeben und konnte ihn erst 1888 nach zwanzigjähriger Dozentur am *Worcester College* wieder einnehmen.

⁵³ Koot (1987, S. 60): "Thorold Rogers added a statistical dimension to the developing discipline of economic history." Heaton (1949, S. 4) bemerkt hierzu nicht ohne Ironie, daß Rogers so viele Fehler begangen habe, daß diese sich gegenseitig aufhoben und seine Ergebnisse damit vermutlich recht genau gewesen seien.

⁵⁴ Siehe oben, Abschnitte A 4 b, B 2 b.

⁵⁵ Cunningham (1892). William Cunningham studierte zunächst Theologie und zeitweise in

später begann die institutionelle Etablierung des Faches. Die erste Dozentur ("full-time lectureship") für Wirtschaftsgeschichte wurde 1904 an der *London University* geschaffen, dann folgten Manchester (1905), Oxford (1907), Edinburgh (1908) und Belfast (1910). Der erste Lehrstuhl für Wirtschaftsgeschichte wurde 1910 in Manchester für G. Unwin eingerichtet, der zweite 1921 in London.⁵⁶ Nach Unwin wurde die englische Wirtschaftsgeschichte vor allem von J. Clapham geprägt.⁵⁷ Als Clapham 1931 die "Wirtschaftsgeschichte als Disziplin" in der *Encyclopedia of the Social Sciences* vorstellte, definierte er sie als "branch of general institutional history", stellte aber zugleich fest, daß es

"[...] would not be difficult to give some simple aspects of institutional history a more quantitative and graphic treatment than they have generally received. [...] Every economic historian should, however, have acquired what might be called the statistical sense, the habit of asking in relation to any institution, policy, group or movement the question: how large? how long? how often? how representative?"⁵⁸

Solche Fragen, die keine elaborierten Methoden erforderten, sondern nur über Tatbestände Klarheit schaffen wollten (eine Aufgabe, die zweifellos nicht zu unterschätzen ist), wurden auch dem Unternehmen des ein Jahr zuvor von der Laura-Spelman-Rockefeller-Stiftung ins Leben gerufenen *Internationalen Wissenschaftlichen Komitees für die Erforschung der Preise* unter Leitung von W. H. Beveridge zugrundegelegt.⁵⁹ Die zum Teil erst Jahrzehnte später publizierten Quellenbände brachten zwar reichhaltige statistische Daten, die meisten Arbeiten beschränkten sich jedoch auf eine Rekonstruktion und Wiedergabe des Zahlenmaterials. Lediglich die zwischen 1934 und 1947 publizierten Untersuchungen des – von J. M. Keynes hochgeschätzten – Harvard-Ökonomen E. J. Hamilton über den Zusammenhang des Preisniveaus und der amerikanischen Silberimporte in Spanien wagten weitergehende Interpretationen des Materials. Auch diese basierten jedoch nicht auf den zur Verfügung stehenden statistischen Methoden, sondern auf visuellen Inspektionen von Zeitreihendarstellungen.⁶⁰

Tübingen bei Vertretern der historischen Schule Nationalökonomie. Seit 1878 lehrte er in Cambridge Wirtschaftsgeschichte. Marshalls Einführung der Partialanalyse beurteilt Hansmeyer (1973, S. 498) treffend: "Der partialanalytisch arbeitende Forscher bedient sich gern des sogenannten Gedankenexperiments auf der Grundlage plausibler oder unmittelbar einsichtiger Annahmen; dieses Vorgehen erwies sich in der Folgezeit als besonders geschickte Methode, widersprechenden Erfahrungstatsachen auszuweichen."

⁵⁶ Vgl. Koot (1987, S. 212). George Unwin, der bei Schmoller in Berlin studiert hatte, wurde 1908 Dozent für Wirtschaftsgeschichte in Edinburgh.

⁵⁷ Sir John Harold Clapham [– 1946], ein Schüler von A. Marshall, war von 1928 bis 1938 Professor für Wirtschaftsgeschichte in Cambridge. Er war Herausgeber der *Cambridge Economic History of Europe* (1941ff). Sein Nachfolger wurde M. M. Postan. Einen guten Eindruck von dem Selbstverständnis englischer Wirtschaftshistoriker geben die von Harte (1971) zusammengestellten Inaugural-Vorlesungen.

⁵⁸ Clapham (1931, S. 327, 328).

⁵⁹ Zu Beveridge siehe oben, Teil B, Anm. 53.

⁶⁰ Zu den Arbeiten des Preiskomitees und ihrer Rezeption siehe Rahlf (1996d, S. 61ff).

Die Entwicklung in den USA

In den Vereinigten Staaten war im Laufe des 19. Jahrhunderts eine Reihe von wirtschaftsgeschichtlichen Untersuchungen erschienen, die aber kaum einer bestimmten Systematik zuzuordnen sind und in der Regel Werke isoliert arbeitender einzelner Personen darstellten.⁶¹ Der weltweit erste Lehrstuhl für Wirtschaftsgeschichte wurde 1892 in Harvard eingerichtet und von W. J. Ashley besetzt.⁶² Ashley veröffentlichte unmittelbar nach seiner Berufung einen programmatischen Beitrag "Über das Studium der Wirtschaftsgeschichte", in dem jedoch statistische Untersuchungen mit keinem Wort erwähnt wurden.⁶³ Auch in der Folgezeit widmete sich die amerikanische Wirtschaftshistoriographie eher qualitativen als quantitativen Themen.⁶⁴ Ashleys Nachfolger in Harvard wurde 1902 E. F. Gay, der sich jedoch zunächst von der Wirtschaftsgeschichte abwandte.⁶⁵ Sein Schüler N. S. B. Gras⁶⁶ beschäftigte sich vorrangig mit "dem Studium der Administration und Geschäftspolitik einzelner Firmen."⁶⁷ Ein weiterer Schüler von Gay, A. Usher, der dessen Lehrstuhl übernahm, war dagegen eher quantitativ orientiert. Usher nahm die ein Jahr zuvor erschienene *Economic History of Modern Britain* von J. Clapham zum Anlaß, die Anwendung der quantitativen Methode auf die Wirtschaftsgeschichte zu propagieren. Clapham hätte in seiner Darstellung zwar mehrfach die Bedeutung quantitativer Angaben betont, aber "despite this use of statistics, the contacts with

⁶¹ Ihnen war jedoch gemein, daß in der Regel quantitative Daten in Zeitreihenform Verwendung fanden. Vgl. zu dieser Entwicklung Cole (1968, S. 557).

⁶² William James Ashley hatte unter anderem bei Karl Knies in Heidelberg studiert und war stark von Schmoller beeinflusst worden. Seit 1888 war er Professor für politische Ökonomie in Toronto. 1901 gründete er an der Universität Birmingham eine Handelshochschule. Nach Gras (1931, S. 326) brachte Ashley "English and German methods and points of view to Harvard".

⁶³ Ashley (1893).

⁶⁴ Siehe dazu Cole (1968, S. 564ff).

⁶⁵ Gay [1867–1946] übernahm die Herausgeberschaft der *New York City Zeitung* und kehrte erst 1926 in die akademische Welt zurück. Er gründete 1920 zusammen mit W. C. Mitchell das *National Bureau of Economic Research* und war für Mitchells zweites Werk *Business Cycles* sein "Hauptberater". Vgl. Mitchell (1927 [1931, S. VII]). In dem erwähnten Komitee zur Preisgeschichte war er für den amerikanischen Teil zuständig. Später übernahm er die Leitung des *Central Bureau of Planning and Statistics*. Den größten Einfluß übte nach eigenen Aussagen Schmoller auf ihn aus. Siehe die instruktive biographische Skizze von Heaton (1949).

⁶⁶ N. S. B. Gras [1884– 1958] besuchte nach dem Studium in Ontario und Harvard verschiedene europäische Universitäten und promovierte über die Entwicklung des englischen Getreidemarktes vom 12. bis zum 18. Jahrhundert. Bis 1918 lehrte er an der *Clark University*, von 1918 bis 1927 als Professor für Wirtschaftsgeschichte an der Universität von Minnesota, schließlich erhielt er den Lehrstuhl für Unternehmensgeschichte an der *Graduate School of Business Administration* in Harvard. Er war Gründer und Herausgeber des *Journal of Economic and Business History*, Herausgeber der *Harvard Studies of Business History* und Präsident der *Business History Foundation*. Anfangs befaßte er sich vor allem mit der Wirtschaftsgeschichte des Mittelalters und sah sich in methodischer/konzeptioneller Hinsicht in der Tradition G. Schmollers. Vgl. Postan (1957).

⁶⁷ Beutin/Kellenbenz (1973, S. 184).

that subject are not developed and we are told that statisticians and historians must, after all, live in separate worlds.”⁶⁸ Nach Usher war die Diskussion um die Anwendung von Methoden in der Wirtschaftsgeschichte zu Unrecht vernachlässigt worden. Seiner Ansicht nach sollte es nun vorrangige Aufgabe sein, “die grundsätzliche Bedeutung der quantitativen Methode für eine systematische Darstellung der Wirtschaftsgeschichte anzuerkennen.”⁶⁹ Vor allem dem Prozeßcharakter wirtschaftshistorischer Erscheinungen sei Rechnung zu tragen, und hierfür habe die Statistik in jüngster Zeit vielversprechende Ansätze entwickelt:

“Better concepts of the process of innovation enable us to explain the nature of the dynamic elements in historical process; the new technique of statistics enables us to describe adequately the process of social growth as a continuously moving equilibrium. This development of statistical technique is an indispensable basis for effective presentation along realistic lines, and there are grounds for believing that the effective application of this new technique to analysis and representation will decisively demonstrate the superiority of the realistic interpretation over idealistic interpretations of history.”⁷⁰

Man könnte hinter dieser Konzeption die von Yule und Slutsky eingeführten stochastischen Prozesse vermuten, doch Usher dachte hier an die Berechnung säkularer Trends.⁷¹ Die Aufgabe des Wirtschaftshistorikers sollte darin bestehen, diese säkularen Trends zu bestimmen und zu kommentieren. Interessanterweise sah er die Möglichkeit einer Arbeitsteilung zwischen Statistikern und Historikern:

“There is some division of labour between the statistician and the historian: the former is more largely interested in the short-run movements, especially the various seasonal and cyclical changes; the latter is primarily concerned with the broad movements over long periods of time.”⁷²

Der Charakter der Arbeitsteilung wurde auch 1938 von C. W. Wright hervorgehoben, als dieser “Natur und Ziele der Wirtschaftsgeschichte” formulierte. Wright ging von der Wirtschaftswissenschaft aus und teilte diese in drei Bereiche ein.⁷³

⁶⁸ Usher (1932, S. 187).

⁶⁹ Vgl. ebda., S. 188. Den Grund sah er vor allem in der unglücklichen Behandlung methodologischer Fragen durch die deutsche historische Schule der Nationalökonomie.

⁷⁰ Ebda., S. 192f. Idealistische Ansätze waren für Usher gleichbedeutend mit “institutionalistischen”. Zu diesen zählte er neben Clapham auch die überwiegende Mehrheit der anglo-amerikanischen und deutschsprachigen wirtschaftshistorischen Arbeiten bis hin zu Sombarts Kapitalismusstudien. Implizit waren hiermit sicher auch die Ansätze von T. Veblen und M. Weber gemeint. Insofern überrascht es, wenn z. B. Boelcke (1987, S. 9) davon ausgeht, daß die Arbeiten von W. C. Mitchell oder S. Kuznets (siehe oben, Kapitel B 2) von Institutionalist^{en} angeregt wurden.

⁷¹ Vgl. Usher (1932, S. 195).

⁷² Ebda., S. 196.

⁷³ Wright (1938). Zur Orientierung: Zu diesem Zeitpunkt waren bereits die Werke von Frisch (1934) und Koopmans (1937) erschienen, die Arbeiten von Tinbergen (1939a), Tinbergen (1939b), Schumpeter (1939) noch nicht, Mitchell (1913), Mitchell (1927) sowie Moore (1914) und Moore (1923) in der Ökonometrie kaum noch beachtet.

1. Die Wissenschaft der Ökonomie, die die Formulierung ökonomischer Gesetze und Prinzipien zum Inhalt habe,
2. die Techniken des
 - a) Rechnungswesens und der
 - b) Statistik sowie
3. das Studium der Wirtschaftsgeschichte im weiteren Sinne, das noch in
 - a) die sogenannte angewandte Ökonomie und
 - b) Wirtschaftsgeschichte im üblichen Sinne

aufgeteilt werde. Hauptunterscheidungsmerkmale der letzten beiden seien vor allem der räumliche und zeitliche Horizont des Gegenstandes sowie die Intention der jeweiligen Untersuchung. Seiner Ansicht nach seien wirtschaftshistorische Untersuchungen bisher aus zweierlei Blickwinkeln heraus geschrieben worden: zum einen von Wissenschaftlern, deren Ausbildung und Standpunkt der allgemeinen Geschichtswissenschaft entspringt, zum anderen von solchen, die aus der Ökonomie stammen. Deren Interesse sei auf eine Untersuchung des Umgangs historischer Gesellschaften mit Knappheitssituationen gerichtet, woraus man sich in der Regel Aufschlüsse für die Bewältigung gegenwärtiger Probleme erhoffte. Diesen Ansatz bezeichnete Wright als funktional.⁷⁴ Das Hauptziel der Wirtschaftsgeschichte sollte ein funktionales sein: Sie sollte zu einer Erhöhung des Lebensstandards durch das Wissen der Vergangenheit und einer darauf aufbauenden Beratung für zukünftige Richtlinien beitragen.⁷⁵ Sicher werde eine solche Aufgabe von Spezialisten der angewandten Ökonomie besser geleistet werden können, doch könne der Wirtschaftshistoriker ihn dabei unterstützen.⁷⁶

⁷⁴ Wright (1938, S. 689). Leider verwies er nicht auf Beispiele solcher Ansätze. Unabhängig von der konkreten Ausgestaltung ging er auf die seiner Ansicht nach für Wirtschaftshistoriker nötige Ausbildung ein. Die Wirtschaftsgeschichte habe sich zwar auf ökonomische Zusammenhänge zu beschränken und sei Bereichen wie der Religion, Ethik oder Lebensphilosophie nachgeordnet. Daraus resultiere aber auf der anderen Seite eine Reihe von Bedingungen: Neben einem (selbstverständlichen) theoretischen Wissen – sowohl der allgemein akzeptierten Theorien als auch des Standes der Diskussion in laufenden Kontroversen – und Kenntnissen in Methoden des Rechnungswesens und der Statistik sei vor allem eine Kenntnis allgemeiner ökonomischer, aber auch sozialer, politischer und naturwissenschaftlicher Entwicklungen notwendig. Kurzum: Neben dem eigentlichen Kern seiner Beschäftigung sei er auf eine Kenntnis der jeweiligen Fachwissenschaften, die die Wirtschaftsgeschichte tangieren, angewiesen. Man fühlt sich hier an die jüngsten Diskussionen um den Standort der Wirtschaftsgeschichte erinnert, in der in mehreren Beiträgen auch von den Wirtschaftshistorikern die Forderung gestellt wird, mehr oder weniger umfassende Kenntnisse zu besitzen, um bestehen zu können.

⁷⁵ Es muß wohl nicht besonders betont werden, daß eine solche Auffassung in diametralem Gegensatz zur deutschsprachigen Wirtschaftshistoriographie stand.

⁷⁶ "The contribution of economic history to the specialist will consist in providing analogous situations from past experience, making clear the general background, indicating the complex relationships between the various specialized lines of economic activity, and suggesting the trends in the evolving economic order as a whole." Wright (1938, S. 693).

Diese Forderung nach allumfassenden Kenntnissen bei gleichzeitiger Beschränkung der eigenen Rolle auf eine Unterstützungstätigkeit spiegelt bereits ein sinkendes Selbstvertrauen der Wirtschaftsgeschichte wider. Zu diesem Zeitpunkt hatte das akademische Interesse an dem Fach (Kursbesuche, Dissertationen) schon merklich nachgelassen. In der ökonomischen Ausbildung wurden wirtschaftshistorische Pflichtveranstaltungen zunehmend durch solche der Ökonometrie ersetzt.⁷⁷ Vielleicht erklärt sich daraus auch die bereits von Usher angedeutete und in der Folgezeit wachsende Kritik an der mangelnden quantitativen Ausrichtung.⁷⁸ In der 1941 erschienenen ersten Ausgabe des *Journal of Economic History* behandelte S. Kuznets die Beziehung zwischen der Statistik und der Wirtschaftsgeschichte. Er mußte feststellen, daß zwischen "statistischen Ökonomen" auf der einen und Wirtschaftshistorikern auf der anderen Seite kaum Austausche stattgefunden hätten, obwohl doch beide dasselbe "Datenmaterial" verwendeten.⁷⁹ Die Aufgabe statistischer Analyse in der Ökonomie beschrieb er hier wie schon sieben Jahre zuvor in seinem Artikel der *Encyclopedia of the Social Sciences* als "search for patterns of temporal change" in säkularen, zyklischen, saisonalen und zufälligen Komponenten. Als weitere Komponente trat hier jedoch ein "measurement of theoretically formulated relations, largely of the short run" hinzu.⁸⁰ Diese Phase der statistischen Analyse gehe von der Annahme aus, daß die in einer ökonomischen Theorie formulierten invarianten Beziehungen in der Realität existierten; sie bestehe hauptsächlich aus der Korrelation von trendbereinigten, zyklischen Zeitreihen. Dabei werde unterstellt, daß die langfristige Beziehung sich nicht ändere (persistent sei) und sich in einer Maßzahl wie dem Korrelationskoeffizienten ausdrücken lasse. Eine solche Vorgehensweise sei von der Suche nach dem Allgemeinen, Typischen motiviert. Der Wirtschaftshistoriker stelle jedoch das Konkrete und Spezifische der betrachteten regionalen oder temporalen Einheit in den Mittelpunkt des Interesses. Darüber hinaus würden in der Wirtschaftsgeschichte häufig qualitative Informationen verwendet und quantitative Daten eher zur Illustration, aber nicht zur Analyse des Typischen und Persistenten herangezogen. Kuznets Ansicht nach sollte die Wirtschaftsgeschichte ebenso einen Beitrag zur ökonomischen Analyse leisten wie andere Studienbereiche der Ökonomie. Dazu müsse sie jedoch in erheblich stärkeren Maße als bisher statistisch orientiert sein:

"I assume that the basic purposes of economic inquiry are to distinguish in the flow and interrelation of economic activities among the common, persistent elements and those that are

⁷⁷ Vgl. Cole (1968, S. 574), Nardinelli/Meiners (1988, S. 545).

⁷⁸ Diese Kritik wurde auch in bezug auf die mangelnde Verwendung theoretischer Modelle geäußert, siehe etwa Johnson (1941) oder Ashton (1946).

⁷⁹ Dieses Argument war kurz zuvor bereits von Heckscher (1939) hervorgehoben worden. Zu diesem Zeitpunkt waren die Arbeiten von Tinbergen und Schumpeter bereits erschienen und letztere von Kuznets einer vernichtenden Kritik unterzogen worden. Siehe oben, Abschnitt B 2 d. Kuznets selbst hatte schon umfangreiche Untersuchungen über langfristige Wirtschaftsentwicklungen vorgelegt.

⁸⁰ Kuznets (1941, S. 27).

diverse and variable; to measure both; to relate them to recognizable factors in operation; and hence finally to lay the foundation for an unbiased understanding of the present, an intelligent consideration of public policy, and a reasonable prognosis of the future. Whatever specific function economic history is to perform in facilitating attainment of this set of basic aims, it should lean upon statistical analysis far more heavily than it has up to the present time. [...] The question is whether or not economic history has actually utilized the results and methods available to it.”⁸¹

Die letzte Frage beantwortete er mit einem klaren Nein. In den letzten drei Jahrzehnten sei jedoch ein statistisches Instrumentarium entwickelt worden, daß auch für die Wirtschaftsgeschichte zukünftig grundlegende Bedeutung haben sollte. Wenn wir die in diesen Zeilen enthaltenen Andeutungen richtig interpretieren, so laufen sie eindeutig auf die in den Abschnitten B 2 b und d beschriebenen, deskriptiv explorativen Verfahren, im wesentlichen auf die Isolierung von Komponenten hinaus. Daß die von ihm 1929 vorgestellten Slutsky-Modelle nicht weiter erwähnt wurden, mag nicht verwundern. Erstaunlich ist jedoch, daß er für die Analyse von Beziehungen zwischen ökonomischen Reihen seinen eigenen Ansatz der “moving correlations” von 1928, den er damals gerade für die Berücksichtigung historischen Wandels vorgeschlagen hatte, nicht mehr erwähnte. Wie schon in seinen eigenen empirischen Studien dominierte hier eindeutig die univariate Analyse. Die zu dieser Zeit bereits aufblühende ökonometrische Entwicklung, insbesondere deren stochastische Sichtweise, wurde nicht rezipiert.⁸² Es sollte jedoch festgehalten werden, daß die Forderung an die Wirtschaftsgeschichte eindeutig ist: Nur wenn sie sich stärker als bisher statistischen Methoden zuwende, könne sie der Ökonomie einen Dienst erweisen.

Die Frage nach der Bedeutung empirischer Forschung für die Wirtschaftsgeschichte und Ökonomie wurde 1949 von Usher noch einmal aufgegriffen. Usher führte nun zwar zahlreiche philosophische, psychologische und naturwissenschaftliche Ansätze an, die einen modernen Empirizismus rechtfertigen und dessen Relevanz für die Wirtschaftsgeschichte hervorheben sollten, seine Bezugnahme auf Ansätze der philosophischen Wahrscheinlichkeitstheorie stehen jedoch vereinzelt dar.⁸³ Insgesamt gesehen war die Disziplin der Wirtschaftsgeschichte auch in den USA eher qualitativ ausgerichtet und quantitativen Ansätzen gegenüber in weiten Teilen ablehnend eingestellt.⁸⁴

⁸¹ Ebda., S. 32f, 35.

⁸² Noch zurückhaltender hinsichtlich der Anwendung statistischer Verfahren argumentierte er in Kuznets (1951).

⁸³ Vgl. Usher (1949, S. 148 sowie S. 155, Anm. 29).

⁸⁴ Fogel (1995, S. 49): “The leading history journals, even in economic history, initially refused to accept articles with complex tables and even after such articles began to be accepted, equations were absolutely forbidden.”

Die 'New Economic History'

Wir waren zu Beginn unserer Untersuchung von der Verwendung quantitativer Methoden in der Wirtschaftsgeschichte und der von R. Whaples konstatierten "cliometric revolution" ausgegangen, die nach Ansicht ihrer Verfechter zu einer "New Economic History" führte und mittlerweile – nach jahrzehntelanger Auseinandersetzung – zumindest die amerikanische Wirtschaftshistoriographie dominiert. Worin besteht nun die Berechtigung zur Verwendung des Adjektivs "new" bzw. was zeichnet die Cliometrie gegenüber der davon abgegrenzten Wirtschaftsgeschichte aus?⁸⁵

Definiert man diesen Ansatz allgemein durch 1. die Praktizierung einer explizit theoriegeleiteten, neoklassisch ausgerichteten wirtschaftshistorischen Forschung sowie 2. die intensive Verwendung von Massendaten und formaler Verfahren zur Überprüfung der Theorien anhand dieser Daten,⁸⁶ so stellt sich hier zunächst die Frage, worin denn überhaupt der Unterschied zum Selbstverständnis der Ökonometrie besteht.

Eine Suche nach expliziteren Definitionen hilft hier nicht weiter. Das Stichwort "cliometrics" des *New Palgrave* definiert den Ansatz als "amalgam of methods",⁸⁷ oder als eine Sammlung von Methoden, "born of the marriage contracted between historical problems and advanced statistical analysis, with economic theory as bridesmaid and the computer as best man",⁸⁸ das *American Heritage Dictionary* als Studium der Geschichte mit fortgeschrittenen mathematischen Methoden der Datenverarbeitung und -analyse.

Wenn das vorherrschende Charakteristikum also die Verwendung bestimmter Methoden ist,⁸⁹ so muß überraschen, daß die Kritik, die die Cliometrie seitens der "traditionellen" Wirtschaftsgeschichte erfahren hat, methodologische Probleme im engeren Sinne überhaupt nicht thematisierte.⁹⁰ Die Diskussionen setzten bei der Frage an, ob eine Anwendung theoretischer Modelle und ihre Überprüfung anhand von einem – oder für einen – konkreten Zeitraum und Ort überhaupt das Erkenntnisziel der Wirtschaftsgeschichte sein sollte und, falls dieser Aspekt nicht völlig verneint wurde, ob die historischen Daten die Voraussetzungen für die Anwendung

⁸⁵ Die Herkunft des Begriffes "cliometrics" läßt sich eindeutig festmachen. Er wurde erstmals in dem Beitrag von Davis/Hughes/Reiter (1960) in Anlehnung an die Bezeichnung einer informellen Gruppe an der *Purdue University* verwendet. Der Ausdruck "New Economic History" stammt von D. North. Vgl. Goldin (1995).

⁸⁶ In diesem Sinne definierten diesen Ansatz – noch unter dem Eindruck der Aufbruchstimmung der damaligen Forschung – auch Wehler (1973b, S. 20) oder Sarrazin (1974, S. 134).

⁸⁷ Floud (1991, S. 452).

⁸⁸ Fogel/Elton (1983, S. 2), zitiert nach Floud (1991, S. 452).

⁸⁹ In diesem Sinne auch Fogel (1995, S. 52): "By the early 1980s *cliometric methods* were so firmly established in certain fields of history that no scholar in these fields could afford to neglect them." Hervorhebung von uns.

⁹⁰ Siehe stellvertretend für viele andere Redlich (1965) oder Fogel/Elton (1983). Ebenso thematisieren die frühen wissenschaftstheoretischen Studien von McClelland (1975) und Sarrazin (1974) andere Aspekte.

elaborierter statistischer Verfahren erfüllten. Die *Methoden selbst* spielten dagegen keine weitere Rolle.

Es ist natürlich nicht möglich, die zahlreichen bislang vorliegenden cliometrischen Untersuchungen einer systematischen Analyse hinsichtlich des in ihnen verwendeten statistischen Instrumentariums zu unterziehen. Wir wollen hier jedoch die – nicht detailliert zu belegende – These aufstellen, daß die Cliometrie in methodischer Hinsicht dem in Abschnitt B 3 d beschriebenen “Paradigma” der Ökonometrie gefolgt ist und damit auch die dort geschilderten Probleme übernommen hat.⁹¹ Wenn man, wie dies schon E. Heckscher 1939 tat, davon ausgeht, daß der Gegenstand der Wirtschaftsgeschichte sich sogar grundsätzlich nicht von dem der Ökonom(etr)ie unterscheidet, dann ist es naheliegend, daß das dort zur Verfügung stehende, Anfang der sechziger Jahre bereits weit entwickelte und kanonisierte Instrumentarium der Ökonometrie kritiklos übernommen wurde, denn zu diesem Zeitpunkt vermittelte die Ökonometrie das geschlossenste Bild ihrer gesamten Entwicklung.⁹²

Vor dem Hintergrund dieser Entwicklung muß es geradezu grotesk erscheinen, daß die allgemein als “Startschuß” der Cliometrie angesehenen Beiträge von A. Conrad und J. Meyer in eine völlig andere Richtung gingen. Die beiden Ökonomen hatten 1957 auf einer gemeinsam von der *Economic History Association* und dem *National Bureau of Economic Research* abgehaltenen Tagung ein Referat über “The Economics of Slavery in the Antebellum South” präsentiert, in dem sie anhand statistischer Verfahren, verstreuter aus der Sekundärliteratur zusammengestellter Daten und eines wirtschaftstheoretischen Modells die These vertraten, daß der Kauf eines Sklaven für einen Sklavenhalter aus den Südstaaten der USA vor dem Bürgerkrieg eine profitable Investition darstellte. Sie lösten mit dieser im folgenden Jahr veröffentlichten Arbeit – sicher nicht nur wegen des “ökonometrischen” Ansatzes – einen Sturm der Entrüstung aus.⁹³ Wir wollen jedoch nicht dieser Diskussion folgen,⁹⁴ sondern ihrem methodologischen Ansatz nachgehen. Dieser wurde ebenfalls 1957 von den Autoren in einem programmatischen Artikel über die Beziehung von Wirtschaftstheorie, statistischer Inferenz und Wirtschaftsgeschichte dargelegt und argumentierte überraschenderweise bayesianisch.⁹⁵

⁹¹ Hierfür spricht nicht zuletzt die Tatsache, daß die Cliometrie als eigenständige Richtung überhaupt erst entstand, nachdem der Antrag einer Gruppe von Gründungsvätern auf Aufnahme in die *Econometric Society* abgelehnt worden war. Vgl. Hughes (1965). Über die Arbeiten informieren bis 1971 Wright (1971), bis 1986 Crafts (1987) sowie Dumke (1986), über jüngere Williamson (1991) und aus europäischer Sicht Dumke (1992).

⁹² Siehe oben.

⁹³ Conrad/Meyer (1958). Die Autoren waren zu dieser Zeit *assistant professors* für Ökonomie in Harvard. Der Ausdruck “Startschuß” oder “Wasserscheide” ist insofern gerechtfertigt, als zum ersten Mal anhand ökonometrischer Methoden historische Erscheinungen untersucht wurden, ohne einen Gegenwartsbezug herzustellen. Vgl. Komlos (1995, S. 405).

⁹⁴ Siehe dazu Conrad/Meyer (1964).

⁹⁵ Conrad/Meyer (1957).

Conrad und Meyer betonten hier zunächst die Bedeutung einer Vorstellung kausaler Ordnungen, die jeder historischen Darstellung zugrundeliegen sollte. Die von einer Reihe von Philosophen (K. Popper, H. Rickert, B. Croce) vorgebrachte Verneinung der Möglichkeit kausaler Erklärungen in der Geschichte beruhe vor allem auf der Ansicht, daß historische Ereignisse einmalig, komplex und nicht quantifizierbar seien.⁹⁶ Sie betonten zu Recht, daß die ökonometrische Modellbildung eine kausale Ordnung vorgibt, die nur für die in dem Modell enthaltenen Variablen Geltung besitzt.⁹⁷ „Kausales Ordnen ist ein operationeller Begriff, der der Einführung von irgendwelchen unsichtbaren Kräften oder inneren Notwendigkeiten nicht bedarf.“ (S. 147). Ebenso sei die Forderung, daß kausale Erklärungen an die prinzipielle Wiederholbarkeit eines Experimentes gebunden, historische Ereignisse jedoch einmalig seien, unzutreffend. Zum einen seien Experimente zunächst auch einmalige Ereignisse, zum anderen wäre dann auch eine Wissenschaft wie die Astronomie zu keiner kausalen Aussage mehr fähig, da sie es mit sich nicht wiederholenden Ereignissen zu tun habe.⁹⁸ Hier setzte nun die bayesianische Argumentation an, die nicht auf der Wiederholbarkeit der Ereignisse aufbaut, sondern Wahrscheinlichkeitsaussagen subjektiv auffaßt:

„Ausdrücklich verbinden die formalen Tests eine wirkliche numerische Wahrscheinlichkeit mit der Richtigkeit der Hypothese im Licht der festgestellten Ergebnisse. Dies führt die Frage der relativen Glaubwürdigkeit in das empirische Verfahren ein und hilft dem Forscher folglich, den Grad an Glauben zu messen – und das ist zumindest ein wirkliches Ordnungskonzept –, der auf die Hypothese gesetzt werden kann. [...] Im Idealfall würde als das beste Verfahren ein solches erscheinen, bei der formale Tests so adaptiert oder verändert werden, daß ein Maximum von a-priori-Informationen in Betracht gezogen wird. Dies führt zugegebenermaßen zu einer im Grunde Bayesianischen Methode des statistischen Schließens.“⁹⁹

Sie sahen es zwar als problematisch an, daß der bayesianische Ansatz in einem „Morast von Subjektivismus“ versinke, wenn A-priori-Wahrscheinlichkeiten¹⁰⁰ nicht unmittelbar gegeben sind, dennoch waren sie zuversichtlich, daß dadurch Richtlinien vorgegeben werden könnten und die Kommunikation wissenschaftlicher Ergebnisse vereinfacht werde.

In der sich an ihren Vortrag anschließenden Diskussion, in der gerade die anwesenden Ökonomen sich gegen die Anwendung ökonometrischer Modelle und statistischer Tests auf historische Daten aussprachen, wurde dieser zentrale Unterschied zum vorherrschenden ökonometrischen Ansatz ebensowenig gesehen wie

⁹⁶ Vgl. Conrad/Meyer (1957 [1973, S. 146]).

⁹⁷ Sie beziehen sich hier auf ein Beispiel von Simon (1957) über die verschiedenen Einflußmöglichkeiten zwischen den Variablen Wetter, Weizen-Erntemenge und Weizenpreis.

⁹⁸ Vgl. Conrad/Meyer (1957 [1973, S. 148f]). Sie stützen sich in diesem Zusammenhang auf die Argumentation von H. Jeffreys. Siehe dazu oben, Abschnitt A 3 f.

⁹⁹ Conrad/Meyer (1957 [1973, S. 160]). Konkrete Beispiele finden sich in Conrad/Meyer (1964).

¹⁰⁰ Hier hat der Übersetzer leider die Bezeichnung „prior notions and probabilities“ durch den mißverständlichen Ausdruck „die zugrundeliegenden Begriffe und Wahrscheinlichkeiten“ übersetzt.

in späteren cliometrischen Arbeiten.¹⁰¹ Die weitere Entwicklung folgte vielmehr dem von der Ökonometrie vorgezeichneten Weg – freilich ohne daß die Ökonometrie hiervon beeinflußt wurde. Bedeutende Ökonometriker brachten, abgesehen von Ausnahmen wie F. M. Fisher oder R. Basman, die die Regel bestätigten, der Wirtschaftsgeschichte kein Interesse mehr entgegen.¹⁰²

¹⁰¹ Das gilt im übrigen auch für Sarrazin (1974, S. 138f), der den bayesianischen Ansatz von Conrad und Meyer zwar darstellte, aber nicht erkannte, daß dieser im Gegensatz zu der von ihm geschilderten Wissenschaftslogik hypothetisch-deduktiver Modelle steht. Die sich an den Vortrag anschließende Diskussion wurde von Kuznets (1957) zusammengefaßt. Bayesianische Ansätze fanden in der Ökonometrie erst mehrere Jahre später fruchtbaren Boden. Siehe oben, Abschnitt B 3 e. Es muß allerdings betont werden, daß in der Argumentation von Conrad und Meyer unterschiedliche Begriffe und Konzepte (sie sprechen von objektiven Tests, signifikanten Unterschieden, dann wieder von Wahrscheinlichkeiten von Hypothesen und einem "Morast von Subjektivismus") nicht immer klar auseinandergehalten werden.

¹⁰² R. Basman, der Erfinder des "Two-Stage-Least-Squares"-Verfahrens, hat dem Historiker die Rolle eines "Testers" für ökonomische Gesetze zugewiesen, blieb dabei jedoch dem Paradigma der klassischen Ökonometrie verhaftet. Siehe Basman (1965). F. M. Fisher, der gegen Liu (1960) die Möglichkeit der Identifikation in Strukturgleichungsmodellen verteidigte, hatte in einem frühen Aufsatz (der unter anderen von J. R. Meyer und A. H. Conrad begutachtet wurde) in einer wissenschaftsphilosophischen Zeitschrift "Über die Analyse der Geschichte und die Interdependenz der Sozialwissenschaften" reflektiert und darin bezeichnenderweise bayesianisch argumentiert. Sein eigener zeitreihenanalytischer Ansatz fand, soweit wir sehen, ebenso wie dieser Aufsatz weder in der Ökonometrie noch in der Wirtschaftsgeschichte Resonanz. Siehe Fisher (1960), Fisher (1962). Die Ausflüge eines Tinbergen in die Wirtschaftsgeschichte blieben qualitativ. Vgl. Tinbergen (1963).

Resümee: Ein Plädoyer für die Geschichte

Als Ziel unserer Untersuchung hatten wir formuliert, die Konzepte, die den in der quantitativen Wirtschaftsgeschichte verwendeten Methoden und Verfahren zugrundeliegen, herauszuarbeiten und darzustellen. Sowohl Verfahren als auch Konzepte werden der quantitativen Wirtschaftsgeschichte vorrangig von der Disziplin der Ökonometrie geliefert, die wiederum Konzepte der Statistik aufgreift und weiterverarbeitet.

Die Lehrbuchliteratur und universitäre Ausbildung der Gesellschaftswissenschaften in quantitativen Methoden vermitteln in der Regel das Bild eines geschlossenen, in sich widerspruchsfreien Systems, in dem bestimmte Gegenstandsvorstellungen als Selbstverständlichkeiten vermittelt werden. Diesem Bild, das von der quantitativen Wirtschaftsgeschichte übernommen worden ist, wurde eine Darstellung der Entstehung und Entwicklung methodologischer Konzepte in der Statistik und Ökonometrie gegenübergestellt, um so eine angemessenere Grundlage für die Beurteilung der Verwendung quantitativer Verfahren schaffen zu können.

Wir hatten uns an dem wissenschaftstheoretischen Ansatz von T. S. Kuhn orientiert und gingen von der Überzeugung aus, daß jede Wissenschaft – und somit auch die Statistik und Ökonometrie – in ihrer gegenwärtigen Verfassung *nur aus ihrer historischen Entwicklung heraus* verstanden werden kann. Ein solcher Ansatz, der für die Entwicklung der Gesellschaftswissenschaften schon oft herangezogen, zum Teil sogar überstrapaziert wurde, fand in bezug auf statistische und ökonometrische Methodologien bislang keine Verwendung. Wir hatten an entsprechender Stelle die Konzepte der Statistik und der Ökonometrie zusammengefaßt und in ihrer inneren Logik kritisch gewürdigt. Diesen Darstellungen folgte eine Übersicht über die Adaption verschiedener Sichtweisen in der Wirtschaftsgeschichte, aus der wir folgende Punkte festhalten wollen:

Überblickt man die historische Entwicklung der Disziplin, so ist sie von Anfang an durch eine ablehnende Haltung gegenüber statistisch-quantitativen Vorgehensweisen gekennzeichnet. Die jüngere historische Schule der Nationalökonomie, erst recht die sich davon abgrenzende, zu Beginn unseres Jahrhunderts entstandene Wirtschaftsgeschichte, orientierte sich an der einer bestimmten Definition des Historismus verpflichteten Geschichtswissenschaft und nicht an der Übernahme statistischer Konzepte in einen sozioökonomischen Kontext durch Edgeworth oder Yule.¹ Jenem Weg folgte, *grosso modo*, auch die Wirtschaftshistorie in England und den USA.

Im weiteren Verlauf und mit wachsendem Druck seitens der mathematisch-theoretisch ausgerichteten Ökonomie wurde der Statistik eine Bedeutung im Rahmen von zusammenfassenden Beschreibungen für Massenerscheinungen zugebilligt. Von Interesse waren Fragen wie: *“how large? how long? how often? how representative?”*. Letztere wurde jedoch nicht im Sinne einer quantitativen Abschätzung von Unsicherheit verstanden, sondern rein deskriptiv. Auch empirisch oder histo-

¹ Man beachte, daß eine der ersten Arbeiten von Yule sogar den Titel *“Notes on the history of pauperism in England and Wales ...”* trug. Siehe Yule (1896a).

risch orientierte Ökonomen wie Mitchell, Schumpeter oder der spätere Kuznets gingen über diesen Ansatz prinzipiell nicht hinaus. Vor diesem Hintergrund ist es kaum überraschend, daß die Ökonometrie, die auf historische Daten angewiesen war, die Wirtschaftsgeschichte als eigenständige Disziplin nicht rezipierte und von dieser auch nicht rezipiert wurde.

Die seit den sechziger Jahren sich rasch entwickelnde *cliometric (r)evolution* übernahm dann mit den Methoden der Ökonometrie auch die damit verbundenen Konzepte, deren Problematik und Fragwürdigkeit wir thematisiert hatten. Ein solcher Weg war aus logischen Gründen ebensowenig zwingend wie die Übernahme der "klassischen" Inferenzstatistik durch die Ökonometrie. Die Arbeiten von Conrad und Meyer, die den Beginn der Cliometrie markieren, argumentierten – unseres Erachtens zu Recht – bayesianisch, doch dies wurde weder von der Cliometrie noch von ihren Kritikern wahrgenommen. Auch hier folgte, um eine Charakterisierung von Menges aufzugreifen, dem Trompetensolo ein eher kläglich-er Chorgesang, der, so mag man hinzufügen, zudem noch eine andere Melodie intonierte. Nur die Macht eines Paradigmas, hier der Neyman-Pearson-Theorie in ihrer ökonomischen Ausgestaltung durch Haavelmo, kann ein solches Phänomen erklären.

Damit wird unseres Erachtens ein zentraler Punkt übersehen: *die Historizität und Entwicklung von Strukturen*. Dies wollen wir zum Anlaß für einige grundsätzliche Überlegungen nehmen.

Fragen wir hierzu zunächst nach der Verortung der Wirtschaftsgeschichte. Wie unterschiedlich auch immer die Aufgaben und Ziele des Faches definiert werden, über einen Punkt dürfte kein Zweifel bestehen: Es ist zwischen der Ökonomie und der traditionellen Geschichtswissenschaft angesiedelt und somit weder auf eine Analyse zeitloser Strukturen noch auf die Darstellung einmaliger Tatsachen beschränkt. Schumpeter sah in seiner Konjunkturzyklen-Untersuchung von 1939 eine Synthese aus einer theoretischen, historischen und statistischen Analyse. Wir wollen uns kein Urteil erlauben, ob das Werk diese Synthese tatsächlich erreicht hat; ihre Einforderung halten wir aber für sinnvoll und notwendig. Die entscheidende Frage ist hier der *Stellenwert* der jeweiligen Elemente zueinander. Auf drei Aspekte wollen wir in diesem Zusammenhang abschließend hinweisen: die Rolle der Deskription, die Beurteilung des Zufalls und das Problem der Modellbildung.

Deskription als eigenständiges Erkenntnisziel

In der Wirtschaftsgeschichte können nicht, wie in den Naturwissenschaften, Experimente unter kontrollierten Randbedingungen durchgeführt werden. Dieses Problem teilt sie mit weiten Bereichen der Ökonomie. Darüber hinaus ist sie in viel stärkerem Maße als diese auf Rekonstruktionen angewiesen. Ist ein konstituierendes Kriterium der Wirtschaftsgeschichte die Beschäftigung mit Vergangenen, so kann sie auf Erhebungsmodalitäten ökonomischer Größen keinen Einfluß mehr nehmen. Die Messung muß sich auf die – oftmals lückenhafte – Überlieferung beschränken. Es ist daher legitim und notwendig, nicht nur nach *Gründen* zu fragen (= Kausalmodelle zu formulieren), sondern zunächst *Tatsachen* festzustellen.

Die Notwendigkeit leuchtet unmittelbar ein, wenn man sich etwa die bis heute bestehende Unsicherheit über die Zahlen der Entwicklung des deutschen Sozialprodukts seit 1850 oder die Ausführungen O. Morgensterns über die Genauigkeit wirtschaftlicher Beobachtungen vergegenwärtigt. Vor diesem Hintergrund kommt neben der Rekonstruktion von Fakten einer eher qualitativen, heuristischen Hypothesenbildung durch eine angemessene Beschreibung, wie sie etwa in den Arbeiten von Mitchell oder den späteren Arbeiten von Kuznets vorgestellt wurde, eine größere Bedeutung zu, als ihr in der gegenwärtigen ökonometrischen und cliometrischen Forschung zugestanden wird. Es ist sicher kein Zufall, daß Kuznets, der sich zunächst der Rezeption und Weiterentwicklung statistischer Methoden in der Ökonomie zugewendet hatte, diesen statistischen Methoden in seinen späteren Arbeiten keine Bedeutung mehr beimaß. In diesem Zusammenhang können unseres Erachtens Elemente der sogenannten "Frankfurter Schule" der Statistik Verwendung finden.

Die Entwicklung der statistischen Methodenlehre hat in Deutschland unter dieser Bezeichnung einen Sonderweg eingeschlagen, der international keine Beachtung fand und mittlerweile auch hierzulande in Vergessenheit geraten ist.² Ziel dieser Schule war es, auf die Besonderheiten der sozialwissenschaftlichen Statistik hinzuweisen. Gerade gegen die – im naturwissenschaftlichen Rahmen entwickelte und auf möglichst optimale Schätzung konstanter Parameter aus unendlichen Grundgesamtheiten ausgerichtete – theoretische Statistik der angelsächsischen Tradition, die zunehmend auch zum Maßstab in den Gesellschaftswissenschaften wurde, erhob sie seit den fünfziger Jahren verstärkt Einwände. Sie kritisierte, daß die statistische Methodik auf einer allen empirischen Wissenschaften gemeinsam zugrundegelegten Wahrscheinlichkeitstheorie aufbaue, für deren Rechtfertigung die Stichprobentheorie herangezogen werde. Für die Sozialwissenschaften erweise sich eine solche Konzeption jedoch als problematisch, denn:

"Alle im Laufe der Zeit sich ändernden wirtschaftlichen und sozialen Erscheinungen – und mit Entwicklungsvorgängen hat es der Volkswirt in der Regel zu tun – sind nämlich *historischen* Charakters, d. h. sie werden unter den gerade obwaltenden Umständen nur ein einziges Mal hervorgebracht."³

In viel stärkerem Maße als in den Naturwissenschaften müsse daher die Problematik der Begriffsbildung, der *Sinnzusammenhang*, beachtet werden:

² Die Frankfurter Schule, die nicht mit der gleichnamigen philosophischen Richtung zu verwechseln ist, wurde von Franz Zizek am Statistischen Seminar der sozial- und wirtschaftswissenschaftlichen Fakultät der Universität Frankfurt/M. begründet und etablierte sich dort als Schule durch Paul Flaskämper, Adolf Blind und Heinrich Hartwig. Ihre Arbeiten wurden zum Teil von H. Grohmann und G. Menges fortgesetzt. Eine – bisher noch nicht vorliegende – systematische Würdigung der Frankfurter Schule kann hier nicht geleistet werden und muß einer späteren Analyse vorbehalten bleiben. Vgl. vorerst Menges (1976), Menges (1980a), Menges (1982) sowie die nicht in jeder Hinsicht überzeugende Arbeit von Ashikaga (1987).

³ Blind (1953, S. 309). Hervorhebung im Original.

“Nicht irgendein mathematisches Kriterium, sondern ihr jeweiliger Bedeutungsgehalt entscheidet über die Brauchbarkeit einer Maßzahl in der sozialwissenschaftlichen Statistik, unsere Maßzahlen sind niemals der ‘sinnfreie’ und daher ‘unverstehbare’ Ausdruck irgendeiner Naturgesetzlichkeit bei optimaler Metrik, sondern jederzeit die bedeutungsgeladenen Symbole eines kulturrelevanten Seins und Geschehens, das wir *nur* im Verstehen zu erkennen vermögen [...]”⁴

A. Blind und H. Hartwig beriefen sich dabei, wie schon ihre Lehrer F. Zizek und P. Flakämper, explizit auf die Philosophen der badischen Schule (vor allem H. Rickert und W. Windelband). Eine solche Betrachtungsweise hat weitreichende Konsequenzen. Es ist demnach prinzipiell unangemessen, von stabilen und in der Zeit gleichbleibenden Verteilungen sozioökonomischer Erscheinungen auszugehen; Konzepte, die darauf aufbauen, müssen zwangsläufig zu Fehlschlüssen führen.

Halten wir fest: Eine verstehende, historische Sozialwissenschaft, wie sie in unterschiedlicher Ausprägung von G. Schmoller oder M. Weber gefordert wurde, läßt sich durchaus mit der Verwendung statistischer Methoden vereinbaren, allerdings mit einer anderen, als der von der Ökonometrie aus der Biometrie und Physik übernommenen Tradition. Sie erfordert statt dessen eine Methodik, die vorrangig auf eine in sinnvolle Maßzahlen zusammenfassende Beschreibung ausgerichtet ist.

Die Rolle des Zufalls

Kommen wir nun zum zweiten Punkt, der Beurteilung des Zufalls. Die Grundlage der Stochastik brachte 1918 der Physiker M. von Smoluchowski auf den Punkt:

“1. Wie ist es möglich, daß sich der Effekt des Zufalls berechnen lasse, daß also *zufällige Ursachen gesetzmäßige Wirkungen* haben? 2. Wie kann der Zufall entstehen, wenn alles Geschehen nur auf regelmäßige Naturgesetze zurückzuführen ist? oder mit anderen Worten: Wie können *gesetzmäßige Ursachen* eine *zufällige Wirkung* haben?”⁵

Die Beantwortung ergibt sich unseres Erachtens aus einer Umkehrung der zweiten Frage: Nur wenn alles Geschehen auf regelmäßige Naturgesetze zurückzuführen ist, kann ein “Zufall” entstehen, der sich in Form von Wahrscheinlichkeiten berechnen läßt, wenn nämlich Ursachenkonstellationen konstant bleiben. Dabei ist es nicht von Interesse, ob diese Verteilung “wesenhaft” zufällig ist oder nur auf unserer Unkenntnis der wahren verursachenden Gründe beruht. Maßgeblich ist allein die Konstanz der Wirkungskräfte. Zufall im Sinne der Stochastik meint also *nicht* einen Zufall im Sinne von Unberechenbarkeit, wie er etwa vom Historismus eines J. G. Droysen oder F. Meinecke verstanden wurde.

Wie sehr sich dieser statistische Zufallsbegriff von einem Zufall in historischem Sinne unterscheidet, belegt sehr deutlich eine Passage von W. Krelle, der als einer der ersten Ökonomen in Deutschland forderte, Konjunkturzyklen stochastisch aufzufassen und zu modellieren. Seiner Meinung nach hätten sich die bisherigen

⁴ Hartwig (1956, S. 263). Hervorhebung im Original.

⁵ Marian von Smoluchowski, Über den Begriff des Zufalls und den Ursprung der Wahrscheinlichkeitsgesetze in der Physik, in: *Die Naturwissenschaften* 6 (1918), S. 253-263, zitiert nach Schneider (1988, S. 80). Hervorhebungen im Original.

Konjunkturtheorien nicht bewährt, weil sie den "Zufall" nicht berücksichtigten.⁶ Dieser Zufallsbegriff sei jedoch nicht falsch zu verstehen:

"Das [Scheitern der bisherigen Konjunkturtheorien, T. R.] legt den Gedanken nahe, dem "Zufall" im Konjunkturgeschehen eine erheblich größere Bedeutung beizumessen, als es die bisherigen Theorien tun, 'Zufall' muß hier aber vom Gesichtspunkt der Konjunkturtheorie aus verstanden werden; historisch sind die gemeinten Ereignisse natürlich absolut nicht zufällig."⁷

Haben wir es hier also mit einem unauflösbaren Gegensatz zu tun, der nur *entweder* eine stochastische *oder* eine historische Betrachtungsweise zuläßt? R. Metz hat im Rahmen seiner Untersuchung über die Modellierung langfristiger Wachstumschwankungen vorgeschlagen, beide Gesichtspunkte miteinander zu verbinden. Er folgert aus der Diskussion um stochastische Trendmodelle, daß diese "in eindrucksvoller Weise die Bedeutung des Zufalls für den langfristigen Verlauf des Sozialprodukts demonstriert" habe.⁸ Seiner Ansicht nach ist es sinnvoll, "große" und "kleine" Zufallsschocks zu unterscheiden, wobei "große" Schocks selten und unregelmäßig (irregulär) auftreten. Diese können nicht durch ein Modell, sondern nur als jeweils einmalige Erscheinungen, die keinem (Wahrscheinlichkeits-) Gesetz folgen, erfaßt werden. Zwischen diesen Ereignissen sei die Struktur des Prozesses dagegen stabil.

Eine solche Sichtweise, die unseres Erachtens einen wichtigen Schritt zu einer angemesseneren Verwendung statistischer Ansätze in der Wirtschaftsgeschichte darstellt, läßt sich weiterführen. In dem von Metz vorgeschlagenen Ansatz treten "Katastrophen" plötzlich auf und üben einen schlagartigen Einfluß aus. Haben sie einen andauernden Einfluß, so geschieht dies ohne Anpassungsphase. Die in diesem Zusammenhang durchgeführte Ausreißeranalyse im Rahmen der ARIMA-Modellierung nach Box und Jenkins geht von einer Strukturkonstanz des datengenerierenden Prozesses während des gesamten betrachteten Zeitraums aus. Die externen Schocks überlagern den Prozeß oder werden wie normale Schocks absorbiert; sie können langandauernde Wirkung haben, verändern jedoch nicht die Struktur des Prozesses.

Hier drängt sich eine physikalische Analogie auf. Eine solche Vorstellung ist vergleichbar mit Einschlägen von Kometen, die – aus Sicht des Planetensystems – zufällig und irregulär sind. Ab einer bestimmten Größe kommt es zu dauerhaften Veränderungen, die aber aufgrund des Wirkens immerwährend gleicher physikalischer Kräfte (Gravitation, Elektromagnetismus, "starke" und "schwache" Kraft, in diesem Falle Gravitation) sich wieder in einem Gleichgewichtszustand einfinden.

⁶ Vgl. Krelle (1959, S. 473). Er verweist in diesem Zusammenhang auf Slutsky, dessen explizite Berücksichtigung von Zufallseinflüssen eine Ausnahme darstelle.

⁷ Ebda. Diese Trennungslinie wird auch noch von Hendry (1995, S. 10) gezogen. Hendry weist darauf hin, daß die Zeitreihen-Ökonometrie mit der Geschichte vieles gemeinsam habe. Bei beiden spiele Zufälligkeit ("contingency") eine Schlüsselrolle, aber: "Such difficulties significantly affect the structure of econometrics, but do not preclude a scientific approach."

⁸ Metz (1995, S. 430).

Eine Vorstellung permanenter Veränderungen erscheint uns dagegen aus prinzipiellen Erwägungen plausibler.⁹ In diesem Zusammenhang muß jedoch die Frage gestellt werden: *Was ist im sozioökonomischen Kontext überhaupt ein Parameter?* "Parameter" haben hier unserer Ansicht nach eine völlig andere Bedeutung als in der Physik oder Biologie. Wenn man für eine längere Periode einen konstanten Parameter unterstellt, so kann dieser eher deskriptiv oder heuristisch als *durchschnittliche Struktur* für diese Periode bzw. als *idealtypische Konstruktion* angesehen werden, nicht aber als Konstante, die sich allenfalls schlagartig von einem Wert auf einen anderen ändert. Im Rahmen einer bayesianischen Konzeption ist erstere Interpretation möglich. Parameter werden hier als Größen angesehen, über die subjektive formale Wahrscheinlichkeitsaussagen getroffen werden, die A-priori-Informationen mit der Information der Daten kombinieren, ohne sich auf eine Stichprobenkonzeption beziehen zu müssen. Dies führt uns zum letzten Punkt, der Modellproblematik.

Modellbildung im bayesianischen Kontext

Die Weichen der Ökonometrie wurden von Physikern gestellt, die sich als "Sozial-Ingenieure" verstanden und sich implizit oder explizit auf Newton beriefen. Die Aufgabe ökonometrischer Analysen wird seither darin gesehen, ein strukturkonstantes System simultaner ökonomischer Beziehungen zu schätzen. Dieses System kann zwar durchaus dynamisch sein, die Dynamik steht jedoch nicht im Vordergrund.

Anders als in der Physik oder Biologie kommt bei der Analyse ökonomischer Daten jedoch dem *Modell* eine zentrale Bedeutung zu. Die Ökonometrie und damit auch die Cliometrie hat Inferenzkonzepte und Methoden der Physik und Biologie übernommen, ohne die Modellunsicherheit zu berücksichtigen. Wir hatten in Abschnitt A 5 und B 4 darauf hingewiesen, daß diesem Aspekt stärker Rechnung getragen werden sollte.

Wenn man stochastische Modelle unterstellt, dann kommt dem Likelihood-Prinzip eine zentrale Bedeutung zu. Ein stochastisches Modell ist durch seine (multidimensionale) Wahrscheinlichkeitsfunktion vollständig charakterisiert. Liegt eine Realisation aus diesem Modell vor, so ist die gesamte Evidenz der Daten über ein spezifisches Modell in der (multidimensionalen) Likelihoodfunktion enthalten. Sie ist somit ein objektives und hinreichendes Maß für die Information, die in den Daten bezüglich der Parameter vorhanden ist; sie gibt Auskunft über die Plausibilität verschiedener Parameterwerte unter der Bedingung einer spezifischen Realisation. In nichtexperimentellen Wissenschaften wie der Wirtschaftsgeschichte ist sie unseres Erachtens die einzig mögliche Inferenzbasis.

Stellt man zusätzlich die Tatsache historischer Entwicklungen, somit eine prinzipielle Variabilität der "Parameter", in Rechnung und bedenkt darüber hinaus den

⁹ L. W. Isaac und L. J. Griffin haben vor wenigen Jahren in zwei Beiträgen fundamentale Kritik an der allgemein angewandten Vorgehensweise bei der zeitreihenanalytischen Untersuchung historischer Prozesse geübt. Siehe Isaac/Griffin (1989) und Griffin/Isaac (1992). Aus konzeptioneller Sicht geht ihr Ansatz jedoch nicht über Kuznets (1928a) hinaus. Siehe allgemein hierzu Rahlf (1996c).

idealtypischen Charakter sozioökonomischer Parameter, erweist sich der von R. Kalman aufgewiesene Weg, hier in bayesianischer Interpretation, als angemessene Lösung, die etwa anhand der Ansätze von Phillips (1995a, 1996) auch das Problem der Modellunsicherheit berücksichtigen kann.¹⁰

Wir wollen somit die vorliegende Untersuchung mit der Forderung nach einer Berücksichtigung von Deskription *und* Inferenz in der hier dargelegten Sichtweise abschließen. In beiderlei Hinsicht sollte schließlich ein unserer Ansicht nach bislang ebenfalls vernachlässigter Aspekt stärkere Beachtung finden: die Bedeutung von graphischen Darstellungen und Techniken. Schon Galton erreichte einen Großteil seiner Einsichten über graphische Veranschaulichungen, und K. Pearsons System der Häufigkeitskurven entstand aus einem Interesse an statistischen Graphiken heraus. Seine ersten Vorlesungen auf dem Gebiet der Statistik waren ausschließlich graphischen Darstellungen gewidmet.¹¹ Die Befürworter traditioneller Komponentenmodelle der Zeitreihenanalyse ziehen ihre Schlüsse vorrangig aus Abbildungen von Zeitreihen bzw. einzelner Komponenten. Differenzengleichungen fanden nicht aus theoretischen Überlegungen Eingang in die statistische Ökonomie, sondern weil die Graphik einer solchen "Zufallsreihe" eine frappierende Ähnlichkeit mit der Abbildung eines konjunkturanzeigenden Index aufwies. Es ist daher erstaunlich, daß erst seit kurzem überhaupt das Bewußtsein einer Problematik besteht (S. 137f). Gerade die explorativ vorgehende historische Zeitreihenanalyse kann unseres Erachtens von einer stärkeren Beachtung dieses Aspektes profitieren.

In dieser Arbeit wurden Konzepte diskutiert und in ihrer historischen Entwicklung dargestellt. Zukünftige Forschungen müßten die Fruchtbarkeit der befürworteten Ansätze anhand konkreter Beispiele und Techniken erst noch belegen, denn: *the proof of the pudding is the eating*. Wir hoffen jedoch, hiermit zumindest eine vielversprechende Sichtweise vorgeschlagen zu haben.

¹⁰ Eine in konzeptioneller Hinsicht vergleichbare, allerdings noch recht spekulative bayesianische Interpretation wurde 1960 von F. M. Fisher vorgestellt, fand jedoch, soweit wir sehen, keine Rezeption. Siehe Fisher (1960).

¹¹ Vgl. Stigler (1986, S. 332, 327).

A1. Illustration des Likelihood-Prinzips

a. Allgemeines

Man kann sich die Idee des Likelihood-Prinzips sehr gut anhand eines graphischen Beispiels veranschaulichen.¹ Man gehe von einem Experiment aus, das zwei Ausgänge habe und dessen Ziel es ist, die Wahrscheinlichkeit der realisierbaren Ereignisse zu bestimmen, z. B. werfe man eine Münze, um zu überprüfen, ob sie "fair" ist. Dieses Experiment führe man n mal durch und bestimme k , die Anzahl der Ereignisse derselben Art, also z. B. "Wappen" und schätze aufgrund dieser Daten die Wahrscheinlichkeit des Ereignisses $Pr(W)$. Dem Experiment liegt offensichtlich eine Binomialverteilung zugrunde:

$$(1) \quad Pr(k; \pi, n) = \binom{n}{k} \pi^k \times (1 - \pi)^{n-k}$$

In der üblichen Vorgehensweise zeichnet man eine Wahrscheinlichkeitsfunktion für ein festes π , trägt also die Wahrscheinlichkeiten für die Realisierungen der möglichen Ereignisse ab. Diese Wahrscheinlichkeitsfunktionen kann man für mehrere π -Werte hintereinander zeichnen, etwa für $\pi = 0, 0.2, 0.4, 0.6, 0.8$ und 1 . Auf der Y -Achse trage man den Wertebereich der Beobachtungen ab, auf der X -Achse die Werte der Dichtefunktionen, die Z -Achse zeigt dann die Wahrscheinlichkeitsdichte, gegeben eine konkrete Beobachtung und eine konkrete Dichtefunktion (Abb. 20).

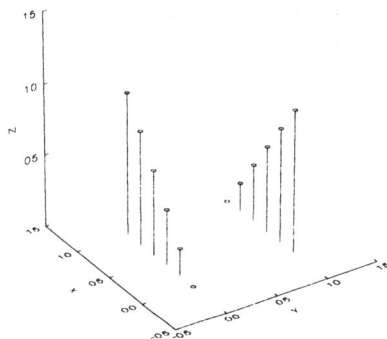


Abb. 20: Wahrscheinlichkeitsfunktionen der Binomialverteilungen mit $\pi = 0, 0.2, 0.4, 0.6, 0.8$ und 1 (aus Richtung der Y -Achse gesehen).

Betrachtet man jetzt nicht die einzelnen Wahrscheinlichkeiten für ein festes π , sondern diejenigen für ein festes X , so erhält man die (sich nicht zu Eins summie-

¹ Vgl. auch die verbalen Beispiele von Stegmüller (1973, S. 114) in Anlehnung an Diehl und Sprott (1965) sowie Menges (1972, S. 281). Ein sehr plastisches Beispiel findet sich bei Schlittgen/Streitberg (1994, S. 269).

renden) Likelihood-Werte. Abb. 21 zeigt zur Verdeutlichung die Likelihood-Werte für $x = 1$ und zum Vergleich die Wahrscheinlichkeitsfunktion für $\pi = 0.2$.

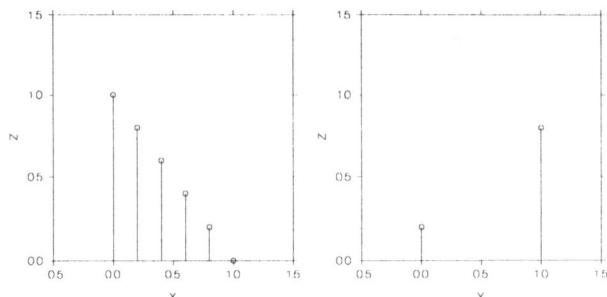


Abb. 21: Likelihood-Werte für $n = 1$, $x = 1$ und $\pi = 0, 0.2, 0.4, 0.6, 0.8, 1$ (links) und Binomialverteilung für $\pi = 0.2$ und $n = 1$ (rechts).

Wir hatten willkürlich sechs verschiedene Werte für π ausgewählt, um die Beziehung zwischen Likelihood und Wahrscheinlichkeit zu demonstrieren. Für die vorliegende Situation ist die Likelihood aber für jeden beliebigen Wert zwischen 0 und 1 definiert. Sie läßt sich einfach bestimmen: Im vorliegenden Fall ist die Likelihoodfunktion mit der Wahrscheinlichkeitsfunktion identisch:

$$(2) \quad L(\pi; k; n) = \binom{n}{k} \pi^k \times (1 - \pi)^{n-k}.$$

Allerdings ist jetzt die Rolle von Parametern und Konstanten vertauscht: k ist fest und π variabel. Für $k = 0$ und $n = 1$ ist

$$(3) \quad L(\pi; 0; 1) = \binom{1}{0} \pi^0 \times (1 - \pi)^{1-0} = (1 - \pi)$$

für $k = 1$ und $n = 1$

$$(4) \quad L(\pi; 1; 1) = \binom{1}{1} \pi^1 \times (1 - \pi)^{1-1} = \pi$$

für $k = 0$ und $n = 2$

$$(5) \quad L(\pi; 0; 2) = \binom{2}{0} \pi^0 \times (1 - \pi)^{2-0} = (1 - \pi)^2$$

usw. Abb. 22 zeigt die vier Likelihoodfunktionen, die sich für $n = 3$ ergeben. Links ist die Likelihoodfunktion für $S = 0$ dargestellt, die mittleren beiden Funktionen geben die "Plausibilitäten" des jeweiligen Parameterwertes bei gegebenem $S = 1$

bzw. $S = 2$ wieder, die rechte schließlich für $S = 3$. Ein Schnitt durch die XY -Ebene an einer beliebigen Stelle π führt zur entsprechenden Wahrscheinlichkeitsverteilung. Für größer werdendes n gleicht sich die Likelihoodfunktion (abgesehen von extremen Realisationen) immer stärker der Normalverteilung an.

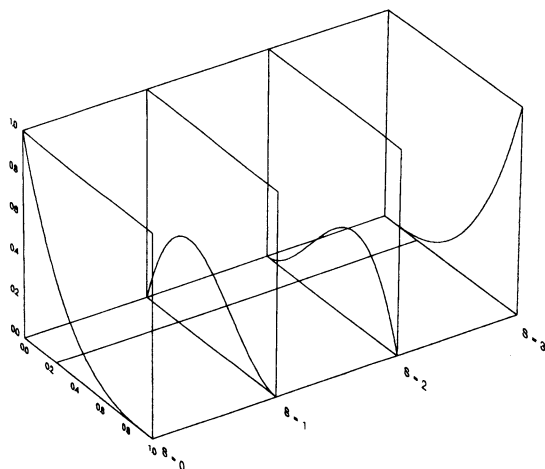


Abb. 22: Gemeinsame Wahrscheinlichkeits- und Likelihoodfunktion für $n = 3$ und $S = 0, S = 1, S = 2, S = 3$.

Abb. 23 zeigt eine schon recht gute Übereinstimmung bei $n = 30$ für $5 < S < 25$.

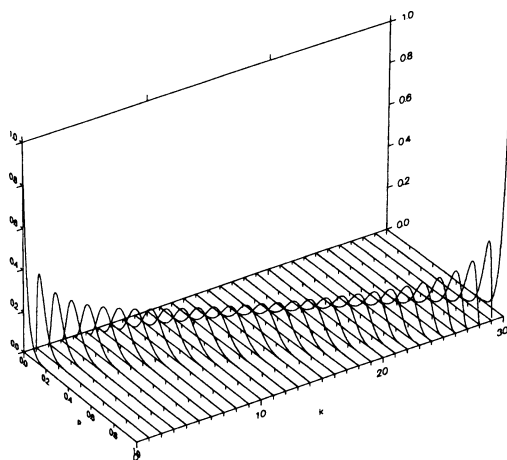


Abb. 23: Gemeinsame Wahrscheinlichkeits- und Likelihoodfunktion für $n = 30$ und $S = 0, S = 1, S = 2, \dots, S = 30$.

Auch für Parameter stetiger Verteilungen existieren analog Likelihoodfunktionen. Die Funktion, die hier am meisten Verwendung findet, ist die Normalverteilung. Diese Funktion ist durch zwei Parameter charakterisiert. Oftmals ist man hier vorrangig an μ und nicht an σ interessiert. Will man nun Aussagen über die Likelihoodfunktion *eines* Parameters machen, so ist dessen *marginale* Likelihoodfunktion zu betrachten. Der oder die anderen Parameter wird / werden dann als “nuisance-parameter” bezeichnet.²

Nachfolgend wollen wir einige Beispiele anführen, um zu illustrieren, in welchen Situationen die Likelihood-Inferenz zu (auch numerisch) anderen Schlüssen gelangt als die klassische häufigkeitstheoretische Inferenz.

1.

Angenommen, es soll die “Null”-Hypothese getestet werden, daß der Anteil der Frauen in einer Grundgesamtheit 0.5 beträgt.³ Wir ziehen eine Stichprobe vom Umfang $n = 10$, bezeichnen mit k die Anzahl der Frauen und führen einen zweiseitigen Test durch. Unter Zugrundelegung der Nullhypothese ($\pi = 0.5$) erhalten wir durch Anwendung der Binomialverteilung die folgenden Wahrscheinlichkeiten:

k	$Pr(S = k)$	k	$Pr(S = k)$
0	1/1024	9	10/1024
1	10/1024	10	1/1024

Wenn wir die Nullhypothese ($\pi = 0.5$) für $k = 0, 1, 9$ oder 10 ablehnen, beträgt das Signifikanzniveau $22/1024 = 0.02$.

Nun werde eine Stichprobe gezogen, und das Resultat laute $k = 9$. Die Nullhypothese wird damit auf einem Niveau von $\theta = 0.02$ abgelehnt. Das Signifikanzniveau besteht hier zum einen aus der Wahrscheinlichkeit für eine Realisation des Ereignisses, das sich ereignet hat, zum anderen aber auch aus Wahrscheinlichkeiten für Ereignisse, *die sich nicht ereignet haben*: “Thus, in classical statistics we are in the strange situation that probabilities of various data that *did not occur* are used as evidence against the null hypothesis.”⁴

² Vgl. etwa Gelman/Carlin/Stern/Rubin (1995, S. 65ff). Hierfür hat sich bislang keine eindeutige Sprachregelung im Deutschen eingebürgert. Schneider (1986, S. 16) verwendet die Bezeichnung “Abfallparameter”.

³ Das folgende Beispiel nach Iversen (1984, S. 10).

⁴ Ebda. Hervorhebungen im Original.

2.

Betrachten wir zwei Wahrscheinlichkeitsverteilungen mit den folgenden Werten:⁵

Tab. 9: Zwei hypothetische Verteilungen $P_0(x)$ und $P_1(x)$.

x	0	1	2	3	4
$P_0(x)$.75	.14	.04	.037	.033
$P_1(x)$.70	.25	.04	.005	.005

Als Teststatistik $T(x)$ wählen wir x und führen zum einen einen Signifikanztest unter Zugrundelegung von Verteilung P_0 durch, zum anderen einen Signifikanztest unter Zugrundelegung von P_1 .⁶ Wir beobachten nun $x = 2$.

Unter Annahme von P_0 beträgt das empirische Signifikanzniveau

$$(6) \quad p_0 = P_0(X \geq 2) = 0.11$$

und unter P_1

$$(7) \quad p_1 = P_1(X \geq 2) = 0.05.$$

In der Logik der Signifikanztests ist es *nicht* das Ziel, P_0 gegen P_1 zu testen, sondern jede Hypothese separat. $x = 2$ gibt nun eine signifikante Evidenz gegen P_0 bei einem 5%-Signifikanzniveau, gegen P_1 jedoch noch nicht einmal auf einem 10%-Signifikanzniveau. Vergleichen wir statt dessen die Likelihoodwerte: Das Verhältnis der Likelihoods (*likelihood ratio*) ist für $x = 2$ gleich Eins. Das Datum $x = 2$ läßt keine Verteilung plausibler erscheinen. Betrachtet man dagegen die Wahrscheinlichkeitsverteilungen isoliert, wie dies in der Signifikanztestkonzeption getan wird, so ist es in der Tat problematisch, daß ein Datum wie $x = 2$ so unterschiedliche Signifikanzniveaus produziert. Dieser Umstand wird von Jeffreys wie folgt formuliert: "[...] a hypothesis which may be true may be rejected because it has not predicted observable results which have not occurred."⁷ Zur Verdeutlichung sei die Situation zusätzlich graphisch veranschaulicht:

⁵ Nach Berger/Wolpert (1988, S. 106) und Cox (1958).

⁶ Dies ist die übliche Vorgehensweise des Signifikanztests: die Überprüfung einer Hypothese anhand *einer* Wahrscheinlichkeitsverteilung im möglichen Stichprobenraum.

⁷ Jeffreys (1961), zitiert nach Berger/Wolpert (1988, S. 106).

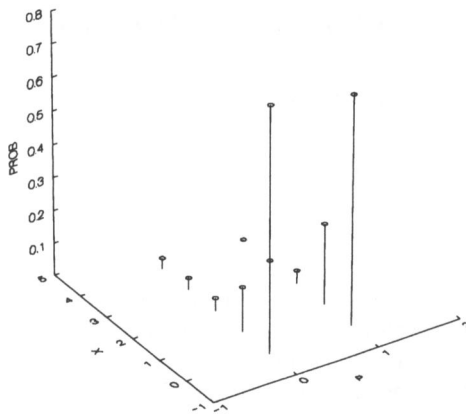


Abb. 24: Gemeinsame Likelihood- und Wahrscheinlichkeitsfunktionen von $P_0(x)$ und $P_1(x)$ aus Beispiel 2. Achsenrichtungen analog zu den vorhergehenden Abbildungen.

Hier wird deutlich, daß die Likelihoodfunktion (Schnitt von links nach rechts bei $X = 2$) beide Resultate gleich bewertet.

3.

Betrachten wir schließlich noch ein letztes Beispiel:⁸ Wir gehen von einem Datum $X = (X_1, X_2, \dots, X_n)$ aus, wobei die X_i i.i.d. $N(\theta, \sigma^2)$ mit bekanntem σ und unbekanntem θ verteilt seien. Die übliche Teststatistik, um $H_0: \theta = \theta_0$ gegen $H_1: \theta \neq \theta_0$ zu testen, ist

$$(8) \quad T(X) = \sqrt{n} \frac{|\bar{X} - \theta_0|}{\sigma},$$

wobei \bar{X} der Mittelwert der Daten ist. Für die Teststatistik $t = T(X)$ ist

$$(9) \quad p = 2(1 - \Phi(t))$$

das Signifikanzniveau mit Standardnormalverteilungsfunktion Φ . Aus der Likelihoodperspektive wird üblicherweise das Likelihoodverhältnis

$$(10) \quad L_{\theta_1} = \frac{f_{\theta_0}(x)}{f_{\theta_1}(x)}$$

⁸ Das folgende Beispiel nach Edwards/Lindman/Savage (1963) aus Berger/Wolpert (1988, S. 107).

betrachtet (s. o.). In dem vorliegenden Fall besteht aber H_1 nicht aus einem einzelnen Wert, sondern ist die zusammengesetzte Hypothese $H_1: \theta \neq \theta_0$. Man kann aber eine untere Grenze angeben:

$$(11) \quad \underline{L} = \frac{f_{\theta_0}(x)}{\limsup_{\theta \neq \theta_0} f_{\theta}(x)}.$$

Wie man zeigen kann, ist in diesem Beispiel

$$(12) \quad \underline{L} = e^{-\frac{1}{2}t^2}.$$

Damit ergeben sich für die vorliegende Situation die folgenden \underline{L} -Werte im Vergleich zu den entsprechenden Signifikanzwerten für verschiedene t -Werte:

Tab. 10: Likelihoodverhältnissgrenzen und Signifikanzniveaus.

t	1.645	1.960	2.576	3.291
p	.100	.050	.001	.001
\underline{L}	.258	.146	.036	.004

Zu $p = 0.05$ ergibt sich ein Wert für \underline{L} von 0.146. Das bedeutet, daß ein Wert der Teststatistik von 1.96 eine Evidenz der Daten von 1 zu 7 gegen H_0 impliziert ("H₁ ist siebenmal plausibler als H₀").

Somit ergeben sich also auch trotz symmetrischer Wahrscheinlichkeitsdichteverteilung und Likelihoodfunktion unterschiedliche Schlußfolgerungen.⁹

b. Anwendung auf Zeitreihen

In Abschnitt A 3 c wurde bereits auf die Arbeit von Barnard/Jenkins/Winsten (1962) hingewiesen. Wir wollen hier zur Illustration die Likelihoodfunktion für einen AR(1)-Prozeß angeben. Für einen AR(1)-Prozeß mit $\mu = 0$ lautet die Likelihoodfunktion:¹⁰

$$(13) \quad L(\theta|x) = \left(2\pi\sigma^2\right)^{-\frac{N}{2}} \left(1-\phi^2\right)^{\frac{1}{2}} e^{\frac{-x_1^2 + x_N^2 + \left(1+\phi^2\right)\sum_{t=2}^{N-1} x_t^2 - 2\phi\sum_{t=1}^{N-1} x_t x_{t+1}}{2\sigma^2}}$$

⁹ Ebda., S. 108 schlagen im folgenden vor, einen Durchschnitt von Likelihoodwerten für die Alternativhypothese zu wählen. Dies führt zu einem Bayes-Ansatz, auf den wir an dieser Stelle nicht näher eingehen wollen.

¹⁰ Vgl. Schlittgen/Streitberg (1994, S. 271). Wir betrachten hier lediglich den einfacheren Fall ohne μ . Falls μ ungleich 0 ist, ist jeweils x_t durch $(x_t - \mu)$ zu ersetzen.

mit $\theta = (\sigma^2, \vartheta)'$. Setzen wir $\sigma^2 = 1$ und betrachten die marginale Likelihoodfunktion für ϑ , die wir so normieren, daß das Maximum gleich Eins ist,

$$(14) \qquad L^*(\vartheta|x) = \frac{L(\vartheta|x)}{\max L(\vartheta|x)},$$

so erhalten wir die folgende Darstellung (Abb. 25). Die linke Hälfte zeigt jeweils einen simulierten AR(1)-Prozeß mit $\vartheta = 0.2$ (oben), $\vartheta = 0.4$ (Mitte) und $\vartheta = 0.9$ (unten), rechts daneben jeweils die relative Likelihoodfunktion.

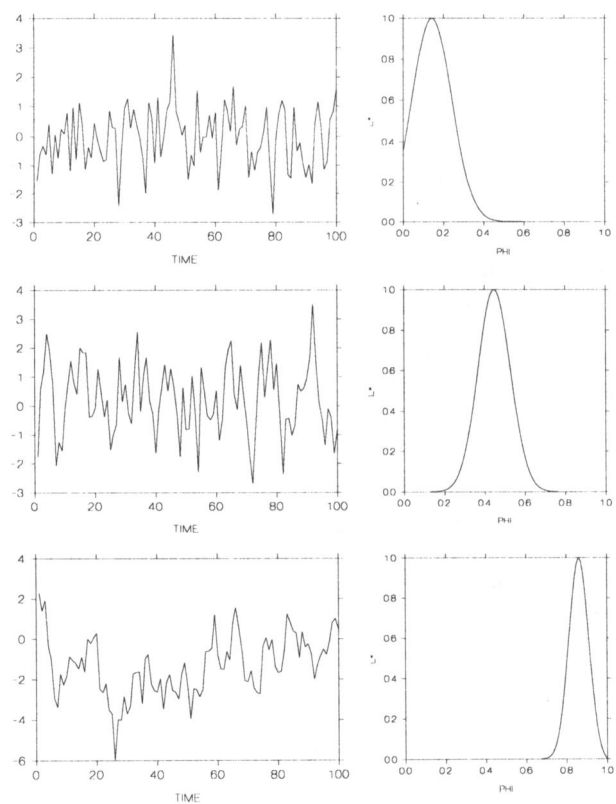


Abb. 25: Relative Likelihoodfunktion für ϕ für drei verschiedene Reihen.

Die Inferenz nach dem Likelihood-Prinzip besteht auch in diesem Zusammenhang darin, nicht nur einen Punkt, sondern die gesamte Likelihoodfunktion in die Inferenzaussage einzubeziehen: "Even if we are using our observation for a decision procedure rather than for an inference problem, we might do better to base our decision on the whole likelihood function rather than on one particular aspect of it such as its maximum."¹¹

Übertragen wir diese Konzeption auf AR(2)-Prozesse, speziell auf die Untersuchung von stochastischen Zyklen,¹² so stehen wir vor folgendem Problem: Eine unmittelbare Betrachtung der Likelihoodfunktion für AR(2)-Prozesse¹³ wäre für die Beurteilung der Zyklicität nicht sinnvoll, da hier lediglich die Plausibilitäten von Parameterkombinationen abgebildet werden. Für eine Beurteilung der relativen Plausibilität von Zyklenlängen benötigen wir aber wiederum eine Funktion dieser beiden Werte. Hierzu wäre es nötig, die Flächen unter der Likelihoodfunktion, die sich über den Parameterkombinationen gleicher Zyklenlänge befinden, miteinander zu vergleichen. Unseres Wissens existieren hierzu aber bislang weder analytische noch heuristische Lösungen.

Wie gezeigt läßt sich das Likelihood-Prinzip unmittelbar in einen Bayes-Ansatz überführen. Nach der Bayesschen Konzeption ist die A-posteriori-Wahrscheinlichkeitsdichte eines Parameters proportional zum Produkt aus der A-priori-Wahrscheinlichkeitsdichte und der Likelihoodfunktion. Wählt man als A-priori-Verteilung eine Gleichverteilung über dem maximalen Intervall, ist dieser Ansatz (numerisch) mit dem Likelihood-Ansatz identisch.

¹¹ Barnard/Jenkins/Winsten (1962, S. 340). Die Basierung der Inferenz auf die Likelihoodfunktion hat darüber hinaus den Vorteil, daß die in der klassischen Inferenzstatistik aufgrund der Einfachheit asymptotischer Aussagen vorgenommene Beschränkung auf stationäre Prozesse hier entfällt: "If we study the likelihood function, no distinction need be drawn between samples from stationary series and samples from wandering [i.e. Random Walks, T. R.] or explosive series. Such distinctions were only drawn because sampling theory depended on asymptotic results, for which rather different analytic methods were required in different cases [...]." Ebda.

¹² Vgl. zu diesem Komplex ausführlich Rahlf (1996b).

¹³ Die Funktion ist in Rahlf (1996b, S. 44, Gleichung 47) angegeben.

A2. Illustration bayesianischer Inferenz

Den Grundgedanken bayesianischer Inferenz wollen wir an drei Beispielen illustrieren.

1.

Gehen wir von einer Person aus, die entweder katholisch oder nicht-katholisch ist. Diese Person könne drei möglichen Populationen entstammen, in denen die Wahrscheinlichkeit, katholisch zu sein, $\pi_1 = 0.3$, $\pi_2 = 0.5$ und $\pi_3 = 0.7$ betrage.¹ Wie groß sind die Wahrscheinlichkeiten, daß Personen Population 1, 2 oder 3 entstammen? Betrachten wir dazu Tab. 11.

Tab. 11: Berechnung der A-posteriori-Wahrscheinlichkeiten für Beispiel 1

Theor. Parameter i	Parameter- wert π_i	Ws. des Par.wertes $Pr(\pi_i)$	Ws. des Datums gg. π_i $Pr(X \pi_i)$	Produkt der Wahrscheinl. $Pr(X \pi_i)Pr(\pi_i)$	A-posteriori- Wahrscheinl. $Pr(\pi_i X)$
(1)	(2)	(3)	(4)	(5) = (4)×(3)	(6) = 5 norm.
1	0.3	1/3	0.3	0.1	0.20
2	0.5	1/3	0.5	0.1667	0.33
3	0.7	1/3	0.7	0.2333	0.47
		$\Sigma = 1$		$\Sigma = 0.5$	$\Sigma = 1.00$

Bevor wir Personen beobachten, halten wir alle drei Parameterwerte für gleich möglich, weisen ihnen eine gleiche A-priori-Wahrscheinlichkeit zu. Damit sich die Einzelwahrscheinlichkeiten zu 1 addieren, beträgt die A-priori-Wahrscheinlichkeit jedes Parameterwertes $Pr(\pi_1) = Pr(\pi_2) = Pr(\pi_3) = 1/3$ (Spalte 3). Nun werde eine Person beobachtet, die katholisch ist. Die Wahrscheinlichkeit einer solchen Beobachtung unter Annahme des jeweiligen Parameterwertes π_i ist natürlich der Parameter selbst (Spalte 4). Die A-posteriori-Wahrscheinlichkeit ergibt sich dann aus dem Produkt der Wahrscheinlichkeit des Parameterwertes und der Likelihood des

¹ Das folgende nach Iversen (1984, S. 13ff) mit z. T. geänderter Notation.

Datums (der Realisation) (Spalte 5), wobei dieser Wert wiederum normiert werden muß, damit die Summe aller (A-posteriori-) Wahrscheinlichkeiten 1 ergibt (Spalte 6).

Die in Tab. 11 ermittelten A-posteriori-Wahrscheinlichkeiten benutzen wir als A-priori-Wahrscheinlichkeiten für eine weitere “Erhebung”, die nun 10 Personen umfasse. Von diesen 10 Personen seien 8 Katholiken. Die Wahrscheinlichkeiten für die Realisation der Daten werden anhand der Binomialverteilung berechnet:

$$(1) \quad Pr(k=8 | \pi = \pi_i) = \binom{10}{8} \pi_i^8 (1 - \pi_i)^{10-8}.$$

Daraus ergeben sich die folgenden A-posteriori-Wahrscheinlichkeiten:

Tab. 12: Berechnung der A-posteriori-Wahrscheinlichkeiten für Beispiel 2

Theor. Parameter <i>i</i>	Parameter- wert π_i	Ws. des Par.wertes $Pr(\pi_i)$	Ws. des Datums gg. π_i $Pr(X \pi_i)$	Produkt der Wahrscheinl. $Pr(X \pi_i) Pr(\pi_i)$	A-posteriori- Wahrscheinl. $Pr(\pi_i X)$
(1)	(2)	(3)	(4)	(5) = (4)×(3)	(6) = (5) norm.
1	0.3	0.20	0.00145	0.0003	0.0024
2	0.5	0.33	0.04394	0.0146	0.1172
3	0.7	0.47	0.23347	0.1097	0.8804
		$\Sigma = 1$		$\Sigma = 0.1246$	$\Sigma = 1.00$

Wir können nun zwei Situationen unterscheiden: 1. Nehmen wir an, es existieren in der Tat drei verschiedene “Populationen” mit Parametern π_i ($i = 1,2,3$). Diese Parameter können entweder theoretische Größen darstellen oder relative Anteile in drei “Grundgesamtheiten”. Das wäre eine “objektive” Bayes-Variante. 2. Andererseits kann es sich aber auch um eine Population handeln, und die verschiedenen Parameterwerte drücken lediglich unser Unwissen über den Wert dieses Parameters aus. Dies wäre eine “subjektive” Bayes-Variante.

Zuletzt noch ein Wort zu der Form der A-priori-Verteilung: Für die A-priori-Verteilung eines Anteilwertes eignet sich die sog. Betaverteilung in besonderem Maße.² Ihre allgemeine Form lautet

² Der Grund für die Wahl dieser Verteilung ist ihre mathematisch leichte Handhabbarkeit. Wir gehen daher nicht weiter auf eine mögliche epistemologische Rechtfertigung für ihre Wahl ein.

$$(2) \quad f(\pi) = C \times \pi^{a-1} (1-\pi)^{b-1}, \quad 0 \leq \pi \leq 1.$$

Die Konstante C dient hier der Normierung der Fläche unter $f(\pi)$ auf 1.³ Für den besonders einfachen Fall von ganzzahligen a und b ergibt sie sich aus

$$(3) \quad C = \frac{(a+b-1)!}{(a-1)!(b-1)!}.$$

Will man die A-priori-Verteilung über übliche Größen wie Mittelwert und Varianz charakterisieren, lassen sich diese, der Einfachheit halber jeweils auf ganze Zahlen gerundet, aus

$$(4) \quad a = \mu \left(\frac{\mu(l-\mu)}{\sigma^2} - 1 \right) \quad \text{und} \quad b = (l-\mu) \left(\frac{\mu(l-\mu)}{\sigma^2} - 1 \right)$$

errechnen. $f(\pi)$ drückt mit zunehmender Potenz von a und b eine immer größer werdende Sicherheit für einen bestimmten Parameter aus.

2.

Wir wollen noch kurz (wiederum anhand eines binomialen Beispiels) illustrieren, daß sich mit zunehmender Evidenz der Daten bei gleichbleibender A-priori-Verteilung die (immer noch subjektive) A-posteriori-Wahrscheinlichkeit der relativen Häufigkeit annähert. Gehen wir dazu von einer β -Verteilung

$$(5) \quad \begin{aligned} f &= C \times \theta^{(l-1)} \times (1-\theta)^{(t-l)} \\ &= C \times (1-\theta)^3 \end{aligned}$$

aus, die dem Wert 0 a priori die höchste, dem Wert 1 die niedrigste Wahrscheinlichkeit zubilligt. Die erste Evidenz der Daten bestehe nun aus zwei Versuchen und einem "günstigen" Resultat: $p = 1/2$. Dafür ergibt sich als Likelihoodfunktion (siehe oben)

$$(6) \quad \begin{aligned} L(\pi; l, 2) &= \binom{2}{1} \pi^1 \times (1-\pi)^{2-1} \\ &= 2\pi(1-\pi) \end{aligned}$$

Aus dem Produkt der A-priori-Verteilung und der Likelihoodfunktion berechnet sich dann die A-posteriori-Wahrscheinlichkeit. Eine zweite Datengrundlage beste-

³ Kolmogoroff-Axiom.

he nun aus 16 Versuchen und 8 “günstigen” Resultaten: damit bleibt $p = 8/16 = 1/2$. Die Likelihoodfunktion

$$(7) \quad L(\pi; 8, 16) = \binom{16}{8} \pi^8 \times (1 - \pi)^{16-8}$$

ist nun jedoch “konzentrierter” um diesen Wert. Daraus ergibt sich ein stärkerer Einfluß auf die A-posteriori-Wahrscheinlichkeit. Die folgenden Abbildungen veranschaulichen beide Situationen.

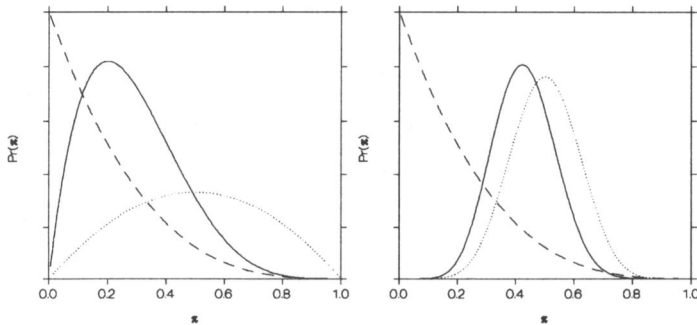


Abb. 26: A-priori-Wahrscheinlichkeit für π (Strichlinie), Likelihoodfunktion (Punktlinie) bei $p=1/2$ (links) und $p=8/16$ (rechts) und A-posteriori-Wahrscheinlichkeit (glatte Linie).

3.

Gehen wir jetzt von einem binomialen Modell aus, für dessen Parameter zwei Hypothesen existieren: Zum einen wird eine Erfolgswahrscheinlichkeit von $\pi = 0.5$ postuliert, zum anderen von $\pi = 0.75$.⁴ Insgesamt werden 20 Realisierungen $X_i \in (0, 1)$, $i = 1, 2, \dots, 20$ betrachtet. Dies ergibt die folgenden Wahrscheinlichkeitsverteilungen:

⁴ Das folgende Beispiel mit leicht geänderter Notation nach Aitkin (1986, S. 109ff).

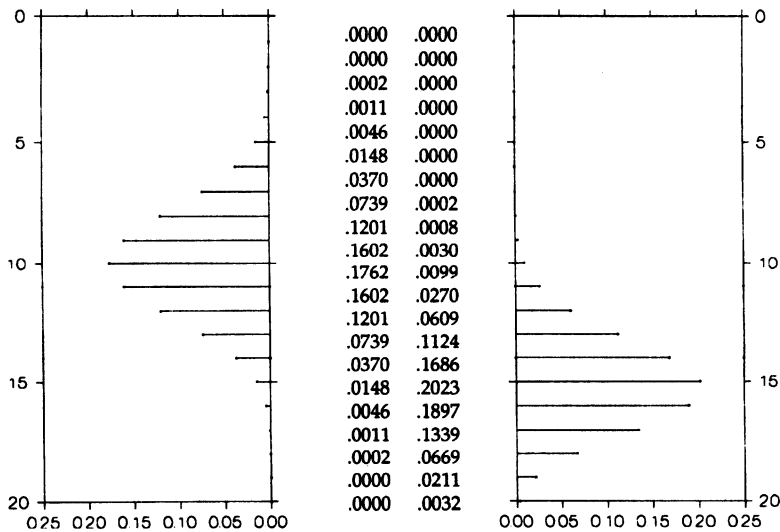


Abb 27: Binomialverteilungen $Pr(S|20, 0.5)$ (links) und $Pr(S|20, 0.75)$ (rechts).

Bei 20 Realisierungen wird beobachtet, daß das Ereignis 14mal eingetreten ist. Die Frage ist nun, wie beide Hypothesen im Lichte dieser Daten zu bewerten sind. In der Logik des Neyman-Pearson-Ansatzes bestehen dazu zwei Möglichkeiten: Zum einen läßt sich $\pi = 0.5$ als Nullhypothese H_0 und $\pi = 0.75$ als Alternativhypothese H_1 formulieren. Der gleichmäßig mächtigste Test führt dann zu einer Ablehnung der Nullhypothese, wenn die Anzahl S der eingetretenen Ereignisse ($X_i = 1$) zu groß ist. Die kritische Region des Tests kann z. B. mit $S \leq 14$ festgelegt werden. Daraus resultiert ein Testniveau von $\alpha = 0.0370 + 0.0148 + 0.0046 + 0.0011 + 0.0002 = 0.0577$ und somit eine Testmacht von $1 - \beta = 0.7857$.⁵ Legt man die kritische Region bei $S \leq 15$ fest, ergibt sich ein Testniveau von $\alpha = 0.0207$ und eine Macht von $1 - \beta = 0.6171$. Wie soll man sich nun entscheiden? Sollte man das Testniveau von 0.05 leicht erhöhen und die Nullhypothese ablehnen, oder sollte man es mit $\alpha = 0.0207$ niedriger ansetzen und die Nullhypothese nicht ablehnen? “Most classical statisticians would feel somewhat uncomfortable in this situation, and might prefer to report $P = .0577$ as the *descriptive significance level* or *P-value* of the test. This leaves the experimenter to make the decision for himself.”⁶

⁵ Ein exaktes Testniveau von $\alpha = 0.05$ läßt sich in diesem diskreten Fall nicht realisieren.

⁶ Aitkin (1986, S. 110). Hervorhebungen im Original.

Zum anderen kann auch $\pi = 0.75$ als Nullhypothese H_0 und $\pi = 0.5$ als Alternativhypothese H_1 formuliert werden. Der gleichmäßig mächtigste Test von H_0 gegen H_1 führt dann zu einer Ablehnung von H_0 , wenn S zu *klein* ist. In diesem Fall ergibt ein $S \leq 11$ ein Testniveau von $\alpha = 0.0410$ bei einer Macht von 0.7483. Die neue Nullhypothese H_0 : $\pi = 0.75$ sollte nur zugunsten von $\pi = 0.5$ abgelehnt werden, falls $S \leq 11$ ist. Das *deskriptive* Signifikanzniveau ist nun erheblich größer:

$\Pr(S \leq 14) = 0.1686 + 0.1124 + 0.0609 + 0.0271 + 0.0099 + 0.0030 + 0.0008 + 0.0002 + \dots = 0.3829$. Demnach würde man H_0 sicher nicht zurückweisen.

Zu welchem Schluß sollte man nun gelangen? Im Rahmen der Neyman-Pearson-Logik kann diese Frage nicht beantwortet werden; eine Asymmetrie zwischen der Null- und Alternativhypothese bleibt bestehen.⁷

Der Bayes-Ansatz bietet nun folgende Lösung an: Jeder Hypothese wird die gleiche A-priori-Wahrscheinlichkeit $p(H_0) = p(H_1) = 0.5$ unterstellt. Dann resultiert für H_0 ($\pi = 0.5$) bei $S = 14$ eine Hypothesenwahrscheinlichkeit von

$$\begin{aligned} (8) \quad \Pr(H_0 | S=14) &= \frac{\Pr(S=14|H_0) \times \Pr(H_0)}{\Pr(S=14|H_0) \times p(H_0) + \Pr(S=14|H_1) \times p(H_1)} \\ &= \frac{0.0370 \times 0.5}{0.0370 \times 0.5 + 0.1686 \times 0.5} \\ &= 0.180 \end{aligned}$$

und analog für H_1 ($\pi = 0.75$) von $\Pr(H_1 | S=14) = (0.1686 \times 0.5) / (0.0370 \times 0.5 + 0.1686 \times 0.5) = 0.820$.

Unter dieser Annahme von gleichen A-priori-Wahrscheinlichkeiten ergibt sich für $p = 0.5$ nach Kenntnis der Daten eine vergleichsweise hohe A-posteriori-Wahrscheinlichkeit. Keine der Hypothesen wird demnach überzeugend gestützt. Es fragt sich hier natürlich, ob eine gleiche A-priori-Wahrscheinlichkeit für beide Parameter sinnvoll ist.⁸ Man kann aber andererseits auch fragen, welche A-priori-Wahrscheinlichkeit zu einer A-posteriori-Wahrscheinlichkeit von 0.0577, dem deskriptiven Signifikanzniveau, führen würde:

$$\begin{aligned} (9) \quad 0.0577 &= \frac{0.0370 \times p(H_0)}{0.0370 \times p(H_0) + 0.1686 \times (1 - p(H_0))} \\ \Rightarrow p(H_0) &= 0.2182. \end{aligned}$$

Das bedeutet, daß man sowohl dem Parameterwert eine A-priori-Wahrscheinlichkeit zuweist, die erheblich unter 0.5 liegt, als auch eine vergleichsweise starke

⁷ Das Problem stellt sich z. B. bei den unit-root-Tests. Vgl. dazu Metz (1995).

⁸ Das Likelihoodverhältnis beider Hypothesen beträgt $L = 0.1686/0.0370 = 4.56$. Damit erscheint H_1 4.56mal plausibler als H_0 .

Evidenz der Daten vorliegen muß, um eine mit dem üblichen Testniveau ($\alpha = 0.05$) vergleichbare A-posteriori-Wahrscheinlichkeit zu erhalten.⁹ Der Grund für diese auf den ersten Blick verblüffende Situation wurde bereits im zweiten Beispiel der Illustration des Likelihood-Prinzips angeführt.¹⁰

Aus Sicht der vorliegenden Daten wird die Evidenz anhand des *P*-Wertes überbewertet: Unbeobachtete, extreme Werte gehen in die Berechnung des *P*-Wertes mit ein.¹¹

⁹ "Vergleichbarkeit" ist hier natürlich nur in numerischem, nicht in epistemologischem Sinne zu verstehen.

¹⁰ Siehe oben, Anhang A1.

¹¹ Eine vergleichbarere Situation würde vorliegen, wenn man lediglich wüßte, daß das Ereignis $X=1$ *mindestens* 14mal eingetreten ist. Dann wäre das Likelihoodverhältnis $L' = 0.7858 / 0.0577 = 13.62$. Gleiche A-priori-Wahrscheinlichkeiten ergeben dann eine A-posteriori-Wahrscheinlichkeit für H_0 von 0.0684, welche in etwa dem *P*-Wert entspricht. Wenn allerdings nur bekannt wäre, daß das Ereignis *höchstens* 14mal eingetreten ist, ergäbe sich ein Likelihoodquotient von $L' = 0.3829 / 0.9793 = 0.391$. Mit gleichen A-priori-Wahrscheinlichkeiten resultiert dann eine A-posteriori-Wahrscheinlichkeit für H_0 von 0.719 und für H_1 von 0.281.

Abbildungen

1.	Bayes "Billardtisch". Nach Menges (1967b, S. 491).	43
2.	Funktionsweise des zweistufigen Galtonschen Brettes. Zeichnung von Karl Pearson, angefertigt nach Vorlagen von Galton. Quelle: Stigler (1986, S. 279).	52
3.	Stadien des iterativen Ansatzes der Modellbildung nach Box/Jenkins. Quelle: Box/Jenkins (1976, S. 19).	101
4.	Struktur der Datenanalyse nach Tukey. Quelle: Tukey (1984, S. xli).	111
5.	"Ausreißer"-Illustration von King (1924, S. 472). Graphisch umgesetzt.	144
6.	Graphische Darstellung der "Kitchin"-Zyklen. Aus: Kitchin (1923, S. 12f).	148
7.	Gleitende Korrelation zwischen Roheisenproduktion und Preisen in fünf Ländern 1865 bis 1903. Quelle: Kuznets (1928a, S. 134).	153
8.	"Quarterly Index of Business Conditions" 1855 bis 1877 und aufsummierte Zufallsgrößen. Quelle: Slutsky (1937, S. 110).	154
9.	Beispiele für aufsummierte Zufallsreihen. Quelle: Working (1934, S. 20).	158
10.	Beispielgraphik für die Darstellung von Zyklensmustern nach Burns und Mitchell. Quelle: Burns/Mitchell (1946, S. 35).	166
11.	Realisation eines VAR(2)-Prozesses aus Gleichung (5).	203
12.	Reaktion von Y_1 , Schock in Y_1 .	204
13.	Reaktion von Y_1 , Schock in Y_2 .	204
14.	Reaktion von Y_2 , Schock in Y_1 .	204
15.	Reaktion von Y_2 , Schock in Y_2 .	204
16.	Logarithmus der Umlaufgeschwindigkeit (v_t) und kurzfristiger Nominalzins (RS_t). Jährliche Daten. Quelle: Hendry/Ericsson (1991, S. 9).	218
17.	Logarithmus der realen Geldmenge ($(m-p)_t$) und des Realeinkommens (i_t). Jährliche Daten. Quelle: Hendry/Ericsson (1991, S. 10).	218
18.	Logarithmus der nominalen Geldmenge (m_t) und des impliziten Preisdeflators (p_t). Jährliche Daten. Quelle: Hendry/Ericsson (1991, S. 11).	218
19.	Aktuelle und angepaßte Werte für $(\bar{m}-\bar{p}-\bar{n})$ nach Gleichung (6). Quelle: Hendry/Ericsson (1991, S. 14).	220
20.	Wahrscheinlichkeitsfunktion der Binomialverteilungen mit $\pi = 0, 0.2, 0.4, 0.6, 0.8$ und 1 (aus Richtung der Y-Achse gesehen).	263
21.	Likelihood-Werte für $n = 1, x = 1$, und $\pi = 0, 0.2, 0.4, 0.6, 0.8$ und 1 (links) und Binomialverteilung für $\pi = 0.2$ und $n = 1$ (rechts).	264
22.	Gemeinsame Wahrscheinlichkeits- und Likelihoodfunktion für $n = 3$ und $S = 0, S = 1, S = 2, S = 3$.	265

23. Gemeinsame Wahrscheinlichkeits- und Likelihoodfunktion für $n = 30$ und $S = 0, S = 1, S = 2, \dots, S = 30$.	265
24. Gemeinsame Likelihood- und Wahrscheinlichkeitsfunktion von $P_0(x)$ und $P_1(x)$ aus Beispiel 2.	268
25. Relative Likelihoodfunktion für ϕ für drei verschiedene Reihen.	270
26. A-priori-Wahrscheinlichkeit für π (Strichlinie), Likelihoodfunktion (Punktlinie) bei $p=1/2$ (links) und $p=8/16$ (rechts) und A-posteriori-Wahrscheinlichkeit (glatte Linie).	275
27. Binomialverteilungen $\Pr(S \mid 20, 0.5)$ (links) und $\Pr(S \mid 20, 0.75)$ (rechts).	276

Tabellen

1. Durchschnittsgröße von Eltern und Kindern nach Galton. Quelle: Stigler (1986, S. 286).	53
2. Wahrscheinlichkeit, anhand von zufälligen Daten "signifikante" Resultate zu erhalten, wenn die zwei besten Regressoren ausgesucht werden. Quelle: Lovell (1983, S. 4).	118
3. Tabellen zur Charakterisierung von Zyklen nach A. F. Burns und W. C. Mitchell. Quelle: Burns/Mitchell (1946, S. 26-32).	163
4. Mögliche erklärende Variablen für die fiktive Entwicklung des Konsums, 1948-1970.	197
5. Konstruierte Modelle, die die artifizielle Konsumreihe generieren. Quelle: Lovell (1983, S. 6).	198
6. Durch "data mining" ausgewählte Variablen. (Auswahl) - Häufigkeiten einzelner Variablen (50 Simulationen).	199
7. In ECONLIT 5/95 erfaßte ökonometrische und bayesianische Publikationen.	206
8. In ECONLIT 5/95 erfaßte ökonometrische und bayesianische Publikationen: Die 10 häufigsten Nennungen.	208
9. Zwei hypothetische Verteilungen $P_0(x)$ und $P_1(x)$.	267
10. Likelihoodverhältnissgrenzen und Signifikanzniveaus.	269
11. Berechnung der A-posteriori-Wahrscheinlichkeiten für Beispiel 1.	272
12. Berechnung der A-posteriori-Wahrscheinlichkeiten für Beispiel 2.	273
13. Synopsis ausgewählter Entwicklungen in der Statistik, Ökonometrie und Wirtschaftsgeschichte.	316

Literaturverzeichnis

Bei abweichenden Seitenangaben im Text beziehen sich diese auf einen angegebenen Wiederabdruck. Titel, die im Text bibliographisch vollständig angegeben wurden (v.a. in Abschnitt A 3 b), werden hier nicht nochmals aufgeführt.

- Abbott, A. (1995), A comment on 'Measuring the agreement between sequences', in: *Sociological Methods & Research* 24, S. 232-243.
- Abel, W. (1974), *Massenarmut und Hungerkrisen im vorindustriellen Europa. Versuch einer Synopsis*, Göttingen.
- Abel, W. (1978), *Agrarkrisen und Agrarkonjunktur. Eine Geschichte der Land- und Ernährungswirtschaft Mitteleuropas seit dem hohen Mittelalter*, 3. Aufl., Hamburg/Berlin. [1. Aufl. 1935]
- Abramovitz, M. (1950), *Inventories in business cycles, with special reference to manufacturers' inventories*, New York (= National Bureau of Economic Research, Studies in Business Cycles 4).
- Achilles, W. (1959), Getreidepreise und Getreidehandelsbeziehungen europäischer Räume im 16. und 17. Jahrhundert, in: *Zeitschrift für Agrargeschichte und Agrarsoziologie* 7, S. 32-55.
- Aitkin, M. (1986), Statistical modelling: the likelihood approach, in: *The Statistician* 35, S. 103-113.
- Aldrich, J. (1995), Correlations genuine and spurious in Pearson and Yule, in: *Statistical Science* 10, S. 364-376.
- Ambrosius, G. et al. (Hgg.) (1996), *Moderne Wirtschaftsgeschichte. Eine Einführung für Historiker und Ökonomen*, München/Wien.
- Anderson, O. D. (1977a), The interpretation of Box-Jenkins time series models, in: *The Statistician* 26, S. 127-145.
- Anderson, O. D. (1977b), Time series analysis and forecasting: another look at the Box-Jenkins approach, in: *The Statistician* 26, S. 285-303.
- Anderson, O. (1914-1915), Nochmals über 'The elimination of spurious correlation due to position in time or space', in: *Biometrika* 10, S. 269-279. [wiederabgedruckt in Kellerer/Mahr/Schneider/Strecker (1963)]
- Anderson, O. (1957), Induktive Logik und statistische Methode, in: *Allgemeines Statistisches Archiv* 41, S. 235-241. [wiederabgedruckt in Kellerer/Mahr/Schneider/Strecker (1963)]
- Anderson, O. (1959), Statistik, (II) Theorie, (2) Mathematische Statistik, in: von Beckerath, E. et al. (Hgg.), *Handwörterbuch der Sozialwissenschaften. Bd. 10*, Stuttgart u. a., S. 46-52. [wiederabgedruckt in Kellerer/Mahr/Schneider/Strecker (1963)]
- Aoki, M. (1987), *State space modelling of time series*, Berlin.
- Artus, P./Laroque, G./Michel, G. (1984), Estimation of a quarterly macroeconomic model with quantity rationing, in: *Econometrica* 52, S. 1387-1414. [wiederabgedruckt in Wallis (1994a)]
- Ashenfelter, O./Pencavel, J. (1969), American trade union growth, 1900-1960, in: *Quarterly Journal of Economics* 83, S. 434-448.

- Ashikaga, S. (1987), Über die Eigenständigkeit der deutschen Sozialstatistik – zur Rolle der Frankfurter Schule in der deutschen Statistik, in: *Jahrbücher für Nationalökonomie und Statistik* 203, S. 456-466.
- Ashley, W. J. (1893), On the study of economic history, in: *Quarterly Journal of Economics* 7, S. 115-136.
- Ashton, T. S. (1946), The relation of economic history to economic theory, in: *Economica* 13, S. 81-96.
- Bailey, R. E./Chambers, M. J. (1993), Long-term demographic interactions in pre-census England, in: *Journal of the Royal Statistical Society A* 156, S. 339-362.
- Baird, D. (1983), The Fisher/Pearson chi-squared controversy: a turning point for inductive inference, in: *British Journal for the Philosophy of Science* 34, S. 109-118.
- Baird, D. (1988), Significance tests, history and logic, in: Kotz, S./Johnson, N. L. (Hgg.), *Encyclopedia of statistical sciences*. Bd. 8, New York u. a., S. 466-472.
- Barlow, R. E. (1992), Introduction to de Finetti (1937) Foresight: its logical laws, its subjective sources, in: Kotz, S./Johnson, N. L. (Hgg.), *Breakthroughs in statistics*. Bd. 1. *Foundations and basic theory*, New York, S. 127-133.
- Barnard, G. A. (1947), The meaning of a significance level, in: *Biometrika* 34, S. 179-182.
- Barnard, G. A. (1949), Statistical inference (with discussion), in: *Journal of the Royal Statistical Society B* 11, S. 115-149.
- Barnard, G. A. (1972a), The unity of statistics (with discussion), in: *Journal of the Royal Statistical Society A* 135, S. 1-14.
- Barnard, G. A. (1972b), Rezension zu: Hacking (1965), in: *British Journal for the Philosophy of Science* 23, S. 123-132.
- Barnard, G. A. (1988), R. A. Fisher – a true Bayesian?, in: *International Statistical Review* 55, S. 183-189.
- Barnard, G. A./Jenkins, G. M./Winsten, C. B. (1962), Likelihood inference and time series (with discussion), in: *Journal of the Royal Statistical Society A* 125, S. 321-375.
- Barnett, V. (1973), *Comparative statistical inference*, New York.
- Bartlett, M. S. (1965), R. A. Fisher and the last fifty years of statistical methodology, in: *Journal of the American Statistical Association* 60, S. 395-409.
- Basmann, R. L. (1965), The role of the economic historian in predictive testing of proffered economic laws, in: *Explorations in Entrepreneurial History (2nd Series)* 2, S. 159-186.
- Bauwens, L./Lubrano, M. (1993), *Trends and breaking points of the Bayesian econometric literature*, Marseille (= G.R.E.Q.E. Document de Travail no 93A06).
- Bauwens, L./Lubrano, M. (1995), Editors' introduction: Bayesian and classical econometric modelling of time series, in: *Journal of Econometrics* 69, S. 1-4.
- Bentzel, R./Wold, H. (1946), On statistical demand analysis from the viewpoint of simultaneous equations, in: *Skandinavisk Aktuarietidskrift* 29, S. 95-113.
- Berger, J. O./Wolpert, R. L. (1988), *The likelihood principle*, 2. Aufl., Hayward (= Institute of Mathematical Statistics. Lecture Notes – Monograph Series 6).
- Bertin, J. (1967), *Semiologie graphique*, Paris.

- Best, H./Schröder, W. H. (1988), Quantitative historische Sozialforschung, in: Meier, C./Rüsen, J. (Hgg.), *Historische Methode*, München (= Beiträge zur Historik 5), S. 235-266.
- Beutin, L./Kellenbenz, H. (1973), *Grundlagen des Studiums der Wirtschaftsgeschichte*, Köln/Wien.
- Beveridge, W. (1922), Wheat prices and rainfall in Western Europe (with discussion), in: *Journal of the Royal Statistical Society A* 85, S. 412-478.
- Bewley, T. F. (Hg.) (1987), *Advances in econometrics: Fifth world congress. Bd. 2*, Cambridge, S. 49-60.
- Birnbaum, A. (1962), On the foundations of statistical inference (with discussion), in: *Journal of the American Statistical Association* 57, S. 269-306. [wiederabgedruckt in Kotz/Johnson (1992a) (ohne Diskussion)]
- Birnbaum, A. (1968), Likelihood, in: Sills, D. L. (Hg.), *International encyclopedia of the social sciences*, New York, S. 299-301.
- Birnbaum, A. (1977), The Neyman-Pearson theory as decision theory and as inference theory: with a criticism of the Lindley-Savage argument for Bayesian theory, in: *Synthese* 36, S. 19-49.
- Bjerkholt, O. (Hg.) (1995), *Foundations of econometrics: the selected essays of Ragnar Frisch*, London.
- Bjornstad, J. F. (1992), Introduction to Birnbaum (1962) On the foundations of statistical inference, in: Kotz, S./Johnson, N. L. (Hgg.), *Breakthroughs in statistics. Bd. 1. Foundations and basic theory*, New York u. a., S. 461-477.
- Blind, A. (1953), Probleme und Eigentümlichkeiten sozialstatistischer Erkenntnis, in: *Allgemeines Statistisches Archiv* 37, S. 301-313.
- Bodkin, R. G./Klein, L. R./Marvah, K. (1991), *A history of macroeconomic model-building*, Aldershot.
- Boelcke, W. A. (1987), *Wirtschafts- und Sozialgeschichte. Einführung, Bibliographie, Problemfelder*, Darmstadt.
- Borchardt, K. (1970), Friedrich Lütge, in: *Jahrbücher für Nationalökonomie und Statistik* 184, S. 1-8.
- Borchardt, K. (1987), Wirtschaftsgeschichte: wirtschaftswissenschaftliches Kernfach, Orchideenfach, Mauerblümchen oder nichts von dem, in: Kellenbenz, H./Pohl, H. (Hgg.), *Historia Socialis et Oeconomica – Festschrift für Wolfgang Zorn zum 65. Geburtstag*, Stuttgart (= VSWG Beiheft 84), S. 17-31.
- Borovcnik, M. (1992), *Stochastik im Wechselspiel von Intuitionen und Mathematik*, Mannheim u. a. (= Lehrbücher und Monographien zur Didaktik der Mathematik 10).
- Bott, D. (1962), Allgemeine und historische Bemerkungen zum Entscheidungsbe-griff, in: *Statistische Hefte* 3, S. 1-38.
- Boumans, M. (1993), Paul Ehrenfest and Jan Tinbergen: a case of limited physics transfer, in: de Marchi, N. (Hg.), *Non-natural social sciences: reflecting on the enterprise of 'More heat than light'*, Durham/London (= Supplement to History of Political Economy 25), S. 131-156.
- Boumans, M. (1995), Frisch on testing of business cycle theories, in: *Journal of Econometrics* 67, S. 129-147.

- Box, G. E. (1976), Science and statistics, in: *Journal of the American Statistical Association* 71, S. 791-799.
- Box, G. E. (1979), Some problems of statistics and everyday life, in: *Journal of the American Statistical Association* 74, S. 1-4.
- Box, G. E. (1980), Sampling and Bayes' inference in scientific modelling (with discussion), in: *Journal of the Royal Statistical Society A* 143, S. 383-430. [wiederabgedruckt in Polson/Tiao (1995a)]
- Box, G. E./Jenkins, G. M. (1973), Some comments on a paper by Chatfield and Prothero and on a review by Kendall, in: *Journal of the Royal Statistical Society A* 125, S. 337-345.
- Box, G. E./Jenkins, G. M. (1976), *Time series analysis, forecasting and control*, 2. Aufl., San Francisco.
- Box, G. E./Tiao, G. C. (1973), *Bayesian inference in statistical analysis*, Reading, Ma.
- Box, G. E./Tiao, G. C. (1975), Some comments on 'Bayes' estimators, in: Fienberg, S. E./Zellner, A. (Hgg.), *Studies in Bayesian econometrics and statistics*, Amsterdam, S. 619-626.
- Braithwaite, R. B. (Hg.) (1931), *The foundation of mathematics and other logical essays by Frank Plumpton Ramsey*, London.
- Braudel, F./Spoonner, F. C. (1967), Prices in Europe from 1450 to 1750, in: Rich, E. E./Wilson, C. H. (Hgg.), *The Cambridge economic history of Europe. Bd. 4. The economy of expanding Europe in the sixteenth and seventeenth centuries*, Cambridge, S. 378-486.
- Breiman, L. (1992), The little bootstrap and other methods for dimensionality selection in regression: X-fixed prediction error, in: *Journal of the American Statistical Association* 87, S. 738-754.
- Brill, H. (1988), Stepwise polynomial regression: royal road or detour, in: *Review* 11, S. 393-411.
- Brillinger, D. R. (1976), Some history of the data analysis of time series in the United States, in: Owen, D. B. (Hg.), *On the history of statistics and probability*, New York/Basel (= Statistics: Textbooks and Monographs 17), S. 267-280.
- Brillinger, D. R. (Hg.) (1984), *The collected works of John W. Tukey. Bd. 1. Time series: 1949-1964. Bd. 2. Time series: 1965-1984*, Belmont.
- Brinkmann, G. (1970), Zur Wissenschaftstheorie der Ökonometrie, in: *Kyklos* 23, S. 205.
- Brinton, W. C. (1914), *Graphical methods for presenting facts*, New York.
- Brown, R. L./Durbin, J./Evans, J. M. (1975), Techniques for testing the constancy of regression relationships over time, in: *Journal of the Royal Statistical Society B* 2, S. 149-163. [wiederabgedruckt in Harvey (1994a)]
- Brush, S. G. (1968), A history of random processes. I. Brownian movement from Brown to Perrin, in: *Archive for the History of the Exact Sciences* 5, S. 1-36.
- Burns, A. F./Mitchell, W. C. (1946), *Measuring business cycles*, New York.
- Capie, F./Mills, T. C. (1992), Money and business cycles in the United States, 1870 to 1913: a reexamination of Friedman and Schwartz, in: *Explorations in Economic History* 29, S. 251-275.
- Carnap, R. (1950), *Logical foundations of probability*, Chicago.

- Carnap, R./Stegmüller, W. (1959), *Induktive Logik und Wahrscheinlichkeit*, Wien.
- Cassel, D./Orear, J. (1971), Applications of statistical inference to physics, in: Godambe, V. P./Sprott, D. A. (Hgg.), *Foundations of statistical inference*, Toronto, S. 280-283.
- Cave-Brown-Cave, F. E. (1904), On the influence of the time factor on the correlation between the barometric heights at stations more than 1000 miles apart, in: *Proceedings of the Royal Society London A* 74, S. 403-411.
- Cave, B. M./Pearson, K. (1914-1915), Numerical illustrations of the variate-difference correlation method, in: *Biometrika* 10, S. 340-355.
- Chatfield, C. (1995), Model uncertainty, data mining and statistical inference (with discussion), in: *Journal of the Royal Statistical Society A* 158, S. 419-466.
- Chatfield, C./Prothero, D. L. (1973), Box-Jenkins seasonal forecasting: problems in a case-study, in: *Journal of the Royal Statistical Society A* 125, S. 295-315.
- Chetty, V. K. (1975), Bayesian analysis of Haavelmo's models, in: Fienberg, S. E./Zellner, A. (Hgg.), *Studies in Bayesian econometrics and statistics in honor of Leonard J. Savage*, Amsterdam (= Contributions to Economic Analysis 86), S. 359-382.
- Christ, C. F. (1983), The founding of the Econometric Society and *Econometrica*, in: *Econometrica* 51, S. 3-6. [wiederabgedruckt in Darnell (1994a)]
- Christ, C. F. (1985), Early progress in estimating quantitative economic relationships in America, in: *American Economic Review* 75, S. 39-52. [wiederabgedruckt in Darnell (1994a)]
- Christ, C. F. (Hg.) (1994a), *Simultaneous equations estimation*, Aldershot (= The International Library of Critical Writings in Econometrics 3).
- Christ, C. F. (1994b), Introduction, in: Christ, C. F. (Hg.), *Simultaneous equations estimation*, Aldershot (= The International Library of Critical Writings in Econometrics 3), S. xi-xxii.
- Cipolla, C. M. (1991), *Between history and economics: an introduction to economic history*, Oxford.
- Claparède, E. (1932), La découverte de l'hypothèse, in: *Journal de Psychologie Normale et Pathologique* 29, S. 648-656. [dt. Übers. in Graumann, C. F. (Hg.) (1969), Denken, 4. Aufl., Köln/Berlin (= Neue Wissenschaftliche Bibliothek 3. Psychologie)]
- Clapham, J. H. (1931), Economic history as a discipline, in: Seligman, E. R./Johnson, A. (Hgg.), *Encyclopedia of the social sciences*. Bd. 5, New York, S. 327-330.
- Cleveland, W. S. (Hg.) (1988a), *The collected works of John W. Tukey*. Bd. 5. *Graphics: 1965-1985*, Belmont.
- Cleveland, W. S. (1988b), Introduction to the graphics volume, in: Cleveland, W. S. (Hg.), *The collected works of John W. Tukey*. Bd. 5. *Graphics: 1965-1985*, Belmont, S. xxxix-xlix.
- Cleveland, W. S. (1993), *Visualizing data*, Summit, New Jersey.
- Cleveland, W. S. (1994), *The elements of graphing data*, 2. Aufl., Summit, New Jersey.
- Cleveland, W. S./McGill, M. E./McGill, R. (1988), The shape parameter of a two-variable graph, in: *Journal of the American Statistical Association* 83, S. 289-300.

- Cochrane, W./Orcutt, G. (1949), Application of least squares regression to relationships containing autocorrelated error terms, in: *Journal of the American Statistical Association* 44, S. 32-61.
- Cole, A. H. (1968), Economic history in the United States: formative years of a discipline, in: *Journal of Economic History* 28, S. 556-589.
- Collins, H./Pinch, T. (1996): *The golem: what everyone should know about science*, Cambridge.
- Conniffe, D. (1992), Keynes on probability and statistical inference and the links to Fisher, in: *Cambridge Journal of Economics* 16, S. 475-489.
- Conrad, A. H./Meyer, J. R. (1957), Economic theory, statistical inference, and economic history, in: *Journal of Economic History* 17, S. 524-544. [wiederabgedruckt in Conrad/Meyer (1964); dt. Übers. in Wehler (1973a)]
- Conrad, A. H./Meyer, J. R. (1958), The economics of slavery in the antebellum south, in: *Journal of Political Economy* 66, S. 95-130.
- Conrad, A. H./Meyer, J. R. (Hgg.) (1964), *The economics of slavery. Studies in econometric history*, Chicago.
- Cooley, T./LeRoy, S. (1985), Atheoretical macroeconometrics. A critique, in: *Journal of Monetary Economics* 16, S. 283-308. [wiederabgedruckt in Poirier (1994a)]
- Cox, D. R. (1958), Some problems connected with statistical inference, in: *Annals of Mathematical Statistics* 29, S. 357-372.
- Cox, D. R. (1990), Role of models in statistical science, in: *Statistical Science* 5, S. 169-174.
- Crafts, N. F. (1987), Cliometrics, 1971-1986: A survey, in: *Journal of Applied Econometrics* 2, S. 171-192.
- Crafts, N. F. (1991), Economic history, in: Eatwell, J./Milgate, M./Newman, P. (Hgg.), *The new Palgrave. A dictionary of economics. Bd. 2, 2. Aufl.*, London/New York/Tokyo, S. 37-41.
- Crafts, N. F./Leybourne, S. J./Mills, T. C. (1989a), Trends and cycles in British industrial production, in: *Journal of the Royal Statistical Society A* 152, S. 43-60.
- Crafts, N. F./Leybourne, S. J./Mills, T. C. (1989b), The climacteric in late Victorian Britain and France: a reappraisal of the evidence, in: *Journal of Applied Econometrics* 4, S. 103-117.
- Crafts, N. F./Leybourne, S. J./Mills, T. C. (1990), Measurement of trend growth in European industrial output before 1914: methodological issues and new estimates, in: *Explorations in Economic History* 27, S. 442-467.
- Crafts, N. F./Leybourne, S. J./Mills, T. C. (1991), Britain, in: Sylla, R./Toniolo, G. (Hgg.), *Patterns of European industrialization: the nineteenth century*, London, S. 109-152.
- Crafts, N. F./Mills, T. C. (1994), Trends in real wages in Britain, 1750-1913, in: *Explorations in Economic History* 31, S. 176-194.
- Craig, L. A./Fisher, D. (1992), Integration of the European business cycle: 1871-1910, in: *Explorations in Economic History* 29, S. 144-168.
- Crum, W. L. (1922), The determination of secular trend, in: *Journal of the American Statistical Association* 18, S. 210-215.
- Crum, W. L. (1923), Cycles of rates on commercial paper, in: *Review of Economic Statistics* 5, S. 17-27.

- Cunningham, W. (1892), The perversion of economic history, in: *Economic Journal* 2, S. 491-506.
- Dale, A. I. (1991), *A history of inverse probability. From Thomas Bayes to Karl Pearson*, New York u. a. (= Studies in the History of Mathematics and Physical Sciences 16).
- Darnell, A. C. (Hg.) (1994a), *The history of econometrics. 2 Bde.*, Aldershot (= The International Library of Critical Writings in Econometrics 4).
- Darnell, A. C. (1994b), Introduction, in: Darnell, A. C. (Hg.), *The history of econometrics. Bd. 1*, Aldershot (= The International Library of Critical Writings in Econometrics 4), S. ix-xx.
- Davidson, R./MacKinnon, J. G. (1993), *Estimation and inference in econometrics*, Oxford/New York.
- Davis, H. T. (1941), *The analysis of economic time series*, Bloomington (= Cowles Commission Monographs 6).
- Davis, L./Hughes, J. R./Reiter, S. (1960), Aspects of quantitative research in economic history, in: *Journal of Economic History* 20, S. 539-547.
- Dawid, A. P. (1983), Inference, statistical: I, in: Kotz, S./Johnson, N. L. (Hgg.), *Encyclopedia of statistical sciences. Bd. 4*, New York u. a., S. 89-105.
- de Finetti, B. (1937), La prévision: Ses lois logiques, ses sources subjectives, in: *Annales des l'Institut Henri Poincaré* 7, S. 1-68. [englische Übersetzung in Kyburg/Smokler (1964), Polson/Tiao (1995a) und teilweise wiederabgedruckt in Kotz/Johnson (1992a)]
- de Finetti, B. (1981), *Wahrscheinlichkeitstheorie. Einführende Synthese mit kritischem Anhang*, Wien/München.
- de Leeuw, J. (1992), Introduction to Akaike (1973) Information theory and an extension of the maximum likelihood principle, in: Kotz, S./Johnson, N. L. (Hgg.), *Breakthroughs in statistics. Bd. 1. Foundations and basic theory*, New York u. a., S. 599-609.
- de Marchi, N. (Hg.) (1988), *The Popperian legacy in economics. Papers presented at a symposium in Amsterdam, December 1985*, Cambridge u. a.
- de Marchi, N./Blaug, M. (Hgg.) (1991), *Appraising economic theories: studies in the methodology of research programs*, Aldershot.
- DeGroot, M. H. (1987), An interview with George Box, in: *Statistical Science* 2, S. 239-258.
- Demmerling, C. (1995), Kuhn, Thomas Samuel, in: Lutz, B. (Hg.), *Metzler Philosophen Lexikon. Von den Vorsokratikern bis zu den Neuen Philosophen*, 2. Aufl., Stuttgart/Weimar, S. 475-480.
- Diehl, H./Sprott, D. A. (1965), Die Likelihoodfunktion und ihre Verwendung beim statistischen Schluß, in: *Statistische Hefte* 6, S. 112-134.
- Dijkstra, T. K. (Hg.) (1988), *On model uncertainty and its statistical implications*, Berlin.
- Doan, T./Litterman, R./Sims, C. A. (1984), Forecasting and conditional projection using realistic prior distributions, in: *Econometric Reviews* 3, S. 1-100.
- Doob, J. L. (1953), *Stochastic processes*, New York.

- Dopsch, A. (1927), Zur Methodologie der Wirtschaftsgeschichte, in: N. N. (Hg.), *Kultur- und Universalgeschichte. Walter Goetz zu seinem 60. Geburtstage dargebracht von Fachgenossen, Freunden und Schülern*, Leipzig/Berlin, S. 518-538.
- Draper, D. (1995), Assessment and propagation of model uncertainty (with discussion), in: *Journal of the Royal Statistical Society B* 57, S. 45-97.
- Drèze, J. H. (1962), *The Bayesian approach to simultaneous equation estimation*, Northwestern University (= ONR Research Memorandum 67).
- Drèze, J. H./Richard, J.-F. (1983), Bayesian analysis of simultaneous equation systems, in: Griliches, Z./Intriligator, M. D. (Hgg.), *Handbook of econometrics*, Amsterdam.
- Drösser, C. (1994), *Fuzzy Logic. Methodische Einführung in krauses Denken*, Reinbeck bei Hamburg.
- Droysen, J. (1863), Die Erhebung der Geschichte zum Rang einer Wissenschaft, in: *Historische Zeitschrift* 9, S. 1-22.
- Dufour, J.-M./Ghysels, E. (Hgg.) (1996a), *Journal of Econometrics. Special issue: Recent developments in the econometrics of structural change*.
- Dufour, J.-M./Ghysels, E. (1996b), Editors' introduction: recent developments in the econometrics of structural change, in: *Journal of Econometrics* 70, S. 1-8.
- Dumke, R. (1986), Clio's Climacteric? Betrachtungen über Stand und Entwicklungstendenzen der Cliometrischen Wirtschaftsgeschichte, in: *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte* 73, S. 457-487.
- Dumke, R. (1992), The future of cliometric history – a european view, in: *Scandinavian Economic History Review* 40, S. 3-28.
- DuMouchel, W. H. (1992), Introduction to Edwards/Lindman/Savage (1963) Bayesian statistical inference for psychological research, in: Kotz, S./Johnson, N. L. (Hgg.), *Breakthroughs in statistics. Bd. 1. Foundations and basic theory*, New York u. a., S. 519-530.
- Durbin, J./Watson, G. S. (1950), Testing for serial correlation in least squares regression, I, in: *Biometrika* 37, S. 409-428. [wiederabgedruckt in Kotz/Johnson (1992a)]
- Durbin, J./Watson, G. S. (1951), Testing for serial correlation in least squares regression, II, in: *Biometrika* 38, S. 159-178. [teilweise wiederabgedruckt in Kotz/Johnson (1992a)]
- Eatwell, J./Milgate, M./Newman, P. (Hgg.) (1990), *The new Palgrave: time series and statistics*, New York/London.
- Edwards, W./Lindman, H./Savage, L. J. (1963), Bayesian statistical inference for psychological research, in: *Psychological Review* 70, S. 193-242. [wiederabgedruckt in Kotz/Johnson (1992a) und Polson/Tiao (1995a)]
- Ehrenberg, A. S./Bound, J. A. (1993), Predictability and prediction (with discussion), in: *Journal of the Royal Statistical Society A* 156, S. 167-206.
- Eisenhart, C. (1983), Laws of error I: Development of the concept. II: The Gaussian distribution. III: Later (non-Gaussian) distributions, in: Kotz, S./Johnson, N. L. (Hgg.), *Encyclopedia of statistical sciences. Bd. 4*, New York u. a., S. 531-566.
- Engle, R. F./Hendry, D. F./Richard, J.-F. (1983), Exogeneity, in: *Econometrica* 51, S. 277-304. [wiederabgedruckt in Hendry (1993), Poirier (1994a) und Harvey (1994a)]

- Epstein, R. J. (1987), *A history of econometrics*, Amsterdam u. a.
- Ezekiel, M. (1928), Statistical analysis and the 'law' of price, in: *Quarterly Journal of Economics* 42, S. 199-227.
- Faraway, J. J. (1992), On the cost of data analysis, in: *Journal of Computational and Graphical Statistics* 1, S. 213-229.
- Feyerabend, P. K. (1970), *Against method. Outline of an anarchistic theory of knowledge*, Minneapolis.
- Fienberg, S. E./Hinkley, D. V. (Hgg.) (1980), *R. A. Fisher: an appreciation*, New York (= Lecture Notes in Statistics 1).
- Fienberg, S. E./Zellner, A. (Hgg.) (1975), *Studies in Bayesian econometrics and statistics in honour of Leonard J. Savage*, Amsterdam (= Contributions to Economic Analysis 86).
- Fisher, F. M. (1960), On the analysis of history and the interdependence of the social sciences, in: *Philosophy of Science* 27, S. 147-158. [wiederabgedruckt in Monz (1991)]
- Fisher, F. M. (1962), *A priori information and time-series analysis*, Amsterdam, North Holland.
- Fisher, R. A. (1922), On the mathematical foundations of theoretical statistics, in: *Philosophical Transactions of the Royal Society of London A* 222, S. 309-368. [wiederabgedruckt in Kotz/Johnson (1992a)]
- Fisher, R. A. (1925), Theory of statistical estimation, in: *Proceedings of the Cambridge Philosophical Society* 22, S. 700-725.
- Fisher, R. A. (1932), Inverse probability and the use of likelihood, in: *Proceedings of the Cambridge Philosophical Society* 28, S. 257-261.
- Fisher, R. A. (1935a), The logic of inductive inference (with discussion), in: *Journal of the Royal Statistical Society* 98, S. 39-82.
- Fisher, R. A. (1935b), The mathematical distributions used in the common tests of significance, in: *Econometrica* 3, S. 353-365.
- Fisher, R. A. (1955), Statistical methods and scientific induction, in: *Journal of the Royal Statistical Society B* 17, S. 69-78.
- Fisher, R. A. (1956), *Statistische Methoden für die Wissenschaft*, 12. Aufl., Edinburgh.
- Fisher, R. A. (1959), *Statistical methods and scientific inference*, London.
- Flaskämper, P. (1949), *Allgemeine Statistik*, 2. Aufl., Hamburg.
- Floud, R. (1980), *Einführung in quantitative Methoden für Historiker*, Stuttgart (= Dt. Bearbeitung hg. von F. Irsigler, engl. 1973, 2. Aufl. 1979).
- Floud, R. (1991), Cliometrics, in: Eatwell, J./Milgate, M./Newman, P. (Hgg.), *The new Palgrave. A dictionary of economics. Bd. 1*, 2. Aufl., London/New York/Tokyo, S. 452-454.
- Fogel, R. W. (1967), The specification problem in economic history, in: *Journal of Economic History* 27, S. 283-308.
- Fogel, R. W. (1989), Afterword: some notes on the scientific methods of Simon Kuznets, in: Kuznets, S. S., *Economic development, the family, and income distribution. Selected essays*, Cambridge u. a. (= Studies in Economic History and Policy), S. 413-438.

- Fogel, R. W. (1995), History with numbers: the American experience, in: Etemad, B./Batou, J./David, T. (Hgg.), *Towards an international economic and social history*, Genf, S. 47-56.
- Fogel, R. W./Elton, G. R. (1983), *Which road to the past? Two views of history*, New Haven/London.
- Fogel, R. W./Engerman, S. L. (1974), *Time on the cross. The economics of American negro slavery. 2 Bde.*, Boston.
- Fraser, D. A. (1983), Inference, statistical: II, in: Kotz, S./Johnson, N. L. (Hgg.), *Encyclopedia of statistical sciences. Bd. 4*, New York u. a., S. 105-114.
- Frazer, W. J./Boland, L. A. (1983), An essay on the foundations of Friedman's methodology, in: *American Economic Review* 73, S. 129-144.
- Frickey, E. (1934), The problem of secular trend, in: *Review of Economic Statistics*.
- Friedman, M. (1940), Review of 'Business cycles in the United States', in: *American Economic Review* 30, S. 657-660.
- Friedman, M. (1948), A monetary and fiscal framework for economic stability, in: *American Economic Review* 38, S. 245-264.
- Friedman, M. (1953), The methodology of positive economics, in: Friedman, M. (Hg.), *Essays in positive economics*, Chicago, S. 3-43.
- Friedman, M./Kuznets, S. S. (1945), *Income from independent professional practice*, New York.
- Friedman, M./Schwartz, A. (1963), *A monetary history of the United States, 1867-1960*, Princeton.
- Friedman, M./Schwartz, A. (1982), *Monetary trends in the United States and the United Kingdom. Their relation to income, prices, and interest rates, 1867-1975*, Chicago.
- Friedman, M./Schwartz, A. (1991), Alternative approaches to analyzing economic data, in: *American Economic Review* 81/1, S. 39-49.
- Frisch, R. (1931), A method of decomposing an empirical series into its cyclical and progressive components, in: *Journal of the American Statistical Association (Supplement)* 26, S. 73-78.
- Frisch, R. (1933a), Propagation problems and impulse problems in dynamic economics, in: N. N. (Hg.), *Economic essays in honour of Gustav Cassel*, London. [wiederabgedruckt in Wallis (1994a), teilweise wiederabgedruckt in Hendry/Morgan (1995a)]
- Frisch, R. (1933b), Editorial, in: *Econometrica* 1, S. 1-4. [wiederabgedruckt in Poirier (1994a)]
- Frisch, R. (1934), *Statistical confluence analysis by means of complete regression systems*, Oslo. [teilweise wiederabgedruckt in Hendry/Morgan (1995a)]
- Frisch, R. (1938), *Autonomy of economic relations. Statistical versus theoretical relations in economic macrodynamics*, Memorandum, Cambridge. [abgedruckt in Hendry/Morgan (1995a)]
- Frisch, R. (1939), A note on errors in time series, in: *Quarterly Journal of Economics* 54, S. 639-640.
- Gandolfo, G. (1990), Eugen Slutsky, in: Eatwell, J./Milgate, M./Newman, P. (Hgg.), *The new Palgrave: time series and statistics*, New York/London, S. 259-262.

- Geisser, S. (1992), Introduction to Fisher (1922) On the mathematical foundations of theoretical statistics, in: Kotz, S./Johnson, N. L. (Hgg.), *Breakthroughs in statistics. Bd. 1. Foundations and basic theory*, New York u. a., S. 1-10.
- Gelman, A./Carlin, J. B./Stern, H. S./Rubin, D. B. (1995), *Bayesian data analysis*, London u. a. (= Texts in Statistical Science).
- Gilbert, C. L. (1989), LSE and the British approach to time series econometrics, in: *Oxford Economic Papers* 41, S. 108-128.
- Godambe, V. P./Sprott, D. A. (Hgg.) (1971), *Foundations of statistical inference*, Toronto.
- Goldin, C. (1995), Cliometrics and the Nobel, in: *Journal of Economic Perspectives* 9, S. 191-208.
- Good, I. J. (1983a), Who is a Bayesian?, in: *American Statistician* 37, S. 95.
- Good, I. J. (1983b), The philosophy of exploratory data analysis, in: *Philosophy of Science* 50, S. 283-295.
- Good, I. J. (1988), Scientific method and statistics, in: Kotz, S./Johnson, N. L. (Hgg.), *Encyclopedia of statistical sciences. Bd. 8*, New York u. a., S. 291-304.
- Good, I. J. (1992), Introduction to Robbins (1955) An empirical Bayes approach to statistics, in: Kotz, S./Johnson, N. L. (Hgg.), *Breakthroughs in statistics. Bd. 1. Foundations and basic theory*, New York u. a., S. 379-387.
- Gosset, W. S. (1908), The probable error of a mean, in: *Biometrika* 6, S. 1-25.
- Gottman, J. M. (1981), *Time-series analysis. A comprehensive introduction for social scientists*, Cambridge u. a.
- Grabas, M. (1991), "Zwangslagen und Handlungsspielräume". Die Wirtschaftsgeschichtsschreibung der DDR im System des real existierenden Sozialismus, in: *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte* 78, S. 501-531.
- Granger, C. W. (1964), *Spectral analysis of economic time series*, Princeton.
- Granger, C. W. (1969), Investigating causal relationships by econometric models and cross-spectral methods, in: *Econometrica* 37, S. 424-438. [wiederabgedruckt in Poirier (1994a) und Harvey (1994a)]
- Granger, C. W. (1980), Testing for causality: a personal viewpoint, in: *Journal of Economics Dynamic and Control* 2, S. 329-352.
- Granger, C. W. (Hg.) (1990a), *Modelling economic series*, Oxford.
- Granger, C. W. (1990b), General introduction: where are the controversies in econometric methodology?, in: Granger, C. W. (Hg.), *Modelling economic series*, Oxford, S. 1-23.
- Granger, C. W./Elliot, C. M. (1967), A fresh look at wheat prices and markets in the eighteenth century, in: *Economic History Review, 2nd ser.* 20, S. 257-265.
- Granger, C. W./Hughes, A. O. (1971), A new look at some old data: the Beveridge wheat price series, in: *Journal of the Royal Statistical Society A* 134, S. 413-428.
- Gras, N. S. (1927), The rise and development of economic history, in: *Economic History Review* 1, S. 12-34.
- Gras, N. S. (1931), Economic history in the United States, in: Seligman, E. R./Johnson, A. (Hgg.), *Encyclopedia of the social sciences. Bd. 5*, New York, S. 325-327.
- Greenstein, B. (1935), Periodogram analysis with special application to business failure in the U.S. 1867-1932, in: *Econometrica* 3, S. 170-198.

- Griffin, L. J./Isaac, L. W. (1992), Recursive regression and the use of 'time' in time-series analysis of historical processes, in: *Historical Methods* 25, S. 166-179.
- Gross, P. R./Levitt, N. (1994), *Higher superstition: the academic left and its quarrels with science*, Baltimore.
- Grüske, K.-D. (Hg.) (1994), *Die Nobelpreisträger der ökonomischen Wissenschaft. Bd. 3. 1989-1993*, Düsseldorf.
- Haase, K. et al. (1992), *MuTi. A menu-driven GAUSS program for multiple time series analysis. Version 1.0*, Kiel.
- Haavelmo, T. (1940), The inadequacy of testing dynamic theory by comparing theoretical solutions and observed cycles, in: *Econometrica* 8, S. 312-321.
- Haavelmo, T. (1943a), Statistical testing of business cycle theories, in: *Review of Economics and Statistics* 25, S. 13-18. [wiederabgedruckt in Darnell (1994a) und Poirier (1994a)]
- Haavelmo, T. (1943b), The statistical implications of a system of simultaneous equations, in: *Econometrica* 11, S. 1-12. [wiederabgedruckt in Hendry/Morgan (1995a)]
- Haavelmo, T. (1944), The probability approach in econometrics, in: *Econometrica (Supplement)* 12, S. 1-118. [wiederabgedruckt in Darnell (1994a), teilweise wiederabgedruckt in Christ (1994a) und Hendry/Morgan (1995a)]
- Haavelmo, T. (1964), *A study in the theory of economic evolution*, Amsterdam (= Contributions to Economic Analysis 3).
- Haavelmo, T. (1994), Ökonometrie und Wohlfahrtsstaat. Nobel-Lesung vom 7. Dezember 1989, in: Grüske, K.-D. (Hg.), *Die Nobelpreisträger der ökonomischen Wissenschaft. Bd. 3. 1989-1993*, Düsseldorf, S. 71-80.
- Hacker, L. M. (1966), The new revolution in economic history, in: *Explorations in Entrepreneurial History* 3, S. 159-175.
- Hacking, I. (1965), *Logic of statistical inference*, Cambridge.
- Hacking, I. (1987), Was there a probabilistic revolution 1800-1930?, in: Krüger, L./Daston, L. J./Heidelberger, M. (Hgg.), *The probabilistic revolution. Bd. 1. Ideas in history*, Cambridge u. a., S. 45-55.
- Hackl, P./Westlund, A. H. (1989), Statistical analysis of 'structural change': an annotated bibliography, in: *Empirical Economics* 14, S. 167-192.
- Hahn, G. J./Meeker, W. Q. (1993), Assumptions for statistical inference, in: *American Statistician* 47, S. 1-11.
- Hall, L. W. (1925), A moving secular trend and moving integration, in: *Journal of the American Statistical Association* 20, S. 13-24.
- Hamilton, J. D. (1994), *Time series analysis*, Princeton.
- Hannan, M./Freeman, J. (1988), The ecology of organizational mortality: American labor unions, 1836-1935, in: *American Journal of Sociology* 94, S. 25-52.
- Hansmeyer, K. H. (1973), Lehr- und Methodengeschichte, in: Ehrlicher, W. et al. (Hgg.), *Kompendium der Volkswirtschaftslehre. Bd. 1, 4. Aufl.*, Göttingen, S. 466-503.
- Hardach, K. W. (1972), Some remarks on German economic historiography, in: *Journal of European Economic History* 1, S. 37-99.
- Harper, W. L./Hooker, C. A. (Hgg.) (1976), *Foundations of probability theory, statistical inference, and statistical theories of science, 2 Bde.*, Reidel, Dordrecht.

- Harte, N. B. (Hg.) (1971), *The study of economic history. Collected inaugural lectures 1893-1970*, London.
- Harter, H. L. (1983), Least squares, in: Kotz, S./Johnson, N. L. (Hgg.), *Encyclopedia of statistical sciences. Bd. 4*, New York u. a., S. 593-598.
- Hartwig, H. (1956), Naturwissenschaftliche und sozialwissenschaftliche Statistik, in: *Zeitschrift für die gesamte Staatswissenschaft* 112, S. 252-266.
- Harvey, A. C. (1981), *The econometric analysis of time series*, London.
- Harvey, A. C. (1987), Applications of the Kalman filter in econometrics, in: Bewley, T. F. (Hg.), *Advances in econometrics: Fifth world congress. Bd. 1*, Cambridge, S. 285-313.
- Harvey, A. C. (1992), *Forecasting, structural time series models and the Kalman filter*, 2. Aufl., Cambridge.
- Harvey, A. C. (Hg.) (1994a), *Time series. 2 Bde.*, Aldershot (= The International Library of Critical Writings in Econometrics 5).
- Harvey, A. C. (1994b), Introduction, in: Harvey, A. C. (Hg.), *Time series. Bd. 1*, Aldershot (= The International Library of Critical Writings in Econometrics 5), S. xi-xviii.
- Hauser, W. (1997), *Die Wurzeln der Wahrscheinlichkeitsrechnung. Die Verbindung von Glücksspieltheorie und statistischer Praxis vor Laplace*, Stuttgart.
- Heaton, H. (1949), The making of an economic historian, in: *Journal of Economic History (Supplement)* 9, S. 1-18.
- Heckman, J. J. (1992), Haavelmo and the birth of modern econometrics: a review of The history of econometric ideas by Mary Morgan, in: *Journal of Economic Literature* 30, S. 876-886.
- Heckscher, E. (1939), Quantitative measurement in economic history, in: *Quarterly Journal of Economics* 53, S. 167-193.
- Heiler, S. (1991), Zeitreihenanalyse – ein kurzer Abriß der Entwicklung, in: *Allgemeines Statistisches Archiv* 75, S. 1-8.
- Heiler, S./Michels, P. (1994), *Deskriptive und Explorative Datenanalyse*, München.
- Hendry, D. F. (1987), Econometric methodology: a personal perspective, in: Bewley, T. F. (Hg.), *Advances in econometrics: Fifth world congress. Bd. 2*, Cambridge, S. 29-48. [wiederabgedruckt in Hendry (1993) und Granger (1990)]
- Hendry, D. F. (1993), *Econometrics – alchemy or science? Essays in econometric methodology*, Oxford.
- Hendry, D. F. (1995), *Dynamic econometrics*, Oxford.
- Hendry, D. F./Ericsson, N. R. (1991), An econometric analysis of U.K. money demand in 'Monetary trends in the United States and the United Kingdom' by Milton Friedman and Anna J. Schwartz, in: *American Economic Review* 81/1, S. 8-38.
- Hendry, D. F./Leamer, E. E./Poirier, D. J. (1990), A conversation of econometric methodology, in: *Econometric Theory* 6, S. 171-261. [wiederabgedruckt in Poirier (1994a)]
- Hendry, D. F./Mizon, G. E. (1985), *Procrustean econometrics: or stretching and squeezing the data*, Tilburg (= Centre for Economic Policy Research, DP No. 68). [wiederabgedruckt in Granger (1990)]
- Hendry, D. F./Morgan, M. S. (1989), A re-analysis of confluence analysis, in: *Oxford Economic Papers* 41, S. 35-52.

- Hendry, D. F./Morgan, M. S. (1994), The ET interview: Professor H. O. A. Wold: 1908-1992, in: *Econometric Theory* 10, S. 419-433.
- Hendry, D. F./Morgan, M. S. (Hgg.) (1995a), *The foundations of econometric analysis*, Cambridge.
- Hendry, D. F./Morgan, M. S. (1995b), Introduction, in: Hendry, D. F./Morgan, M. S. (Hgg.), *The foundations of econometric analysis*, Cambridge, S. 1-82.
- Hendry, D. F./Richard, J.-F. (1983), The econometric analysis of time series (with discussion), in: *International Statistical Review* 51, S. 111-163. [wiederabgedruckt in Hendry (1993)]
- Hill, B. M. (1990), A theory of Bayesian data analysis, in: Geisser, S./Hodges, J. S./Press, S. J./Zellner, A. (Hgg.), *Bayesian and likelihood methods in statistics and econometrics. Essays in honor of George A. Barnard*, New York u. a. (= Studies in Bayesian Econometrics and Statistics 7), S. 49-73.
- Hjorth, J. S. (1994), *Computer intensive statistical methods: validation model selection and bootstrap*, London.
- Hoaglin, D. C. (1985), Exploratory data analysis, in: Kotz, S./Johnson, N. L. (Hgg.), *Encyclopedia of statistical sciences. Bd. 2*, New York u. a., S. 579-583.
- Hodges, J. S. (1990), Can/may Bayesians do pure tests of significance?, in: Geisser, S./Hodges, J. S./Press S. J./Zellner, A. (Hgg.), *Bayesian and likelihood methods in statistics and econometrics. Essays in honor of George A. Barnard*, New York u. a. (= Studies in Bayesian Econometrics and Statistics 7), S. 75-90.
- Hoffman, P./Postel-Vinay, G./Rosenthal, J.-L. (1994), Économie et politique. Les marchés du crédit à Paris, 1750-1840, in: *Annales. Economies – Sociétés – Civilisations* 49, S. 65-97.
- Hooker, R. H. (1901), Correlation of the marriage-rate with trade, in: *Journal of the Royal Statistical Society* 64, S. 485-492. [wiederabgedruckt in Hendry/Morgan (1995a)]
- Hooker, R. H. (1905), On the correlation of successive observations; illustrated by corn prices, in: *Journal of the Royal Statistical Society* 68, S. 696-703.
- Horwich, P. (Hg.) (1993), *World changes. Thomas Kuhn and the nature of science*, Cambridge, Mass.
- Hotelling, H. (1934), Analysis and correlation of time series, in: *Econometrica* 2, S. 211.
- Howson, C. (1995), Theories of probability, in: *British Journal for the Philosophy of Science* 46, S. 1-32.
- Hoyningen-Huene, P. (1989), *Die Wissenschaftsphilosophie Thomas Kuhns. Rekonstruktion und Grundlagenprobleme*, Braunschweig.
- Hughes, J. R. (1965), A note in defense of Clio, in: *Explorations in Entrepreneurial History* 3, S. 154.
- Hylleberg, S./Paldam, M. (1991), New approaches to empirical macroeconomics: Editors' introduction, in: *Scandinavian Journal of Economics* 93, S. 121-128.
- Iggers, G. G. (1996), *Geschichtswissenschaft im 20. Jahrhundert. Ein kritischer Überblick im internationalen Zusammenhang*, 2. Aufl., Göttingen.
- Irsigler, F./Metz, R. (1984), The statistical evidence of "long waves" in pre-industrial and industrial times, in: *Social Science Information* 23, S. 381-410.

- Irwin, J. O. (1963), Obituary Sir Ronald Aylmer Fisher, 1890-1962: introduction, in: *Journal of the Royal Statistical Society A* 126, S. 159-162.
- Isaac, L. W./Griffin, L. J. (1989), Ahistoricism in time-series analysis of historical processes: critique, redirection, and illustrations from U.S. labor history, in: *American Sociological Review* 54, S. 873-890.
- Iversen, G. R. (1984), *Bayesian statistical inference*, Newbury Park u. a.
- Jarausch, K. H./Arminger, G./Thaller, M. (1985), *Quantitative Methoden in der Geschichtswissenschaft*, Darmstadt.
- Jarausch, K. H./Hardy, K. A. (1991), *Quantitative methods for historians*, Chapel Hill/London.
- Jaynes, E. T. (1996), *Probability theory: the logic of science*, Cambridge.
- Jeanne, O. (1995), Monetary policy in England 1893-1914: a structural VAR analysis, in: *Explorations in Economic History* 32, S. 302-326.
- Jeffreys, H. (1939), *Theory of probability*, London/New York.
- Jeffreys, H. (1961), *Theory of probability*, 3. Aufl., Oxford.
- Jenkins, G. M./Watts, D. G. (1968), *Spectral analysis and its applications*, San Francisco.
- Johnson, D. R. (1992), International interest rate linkages and the exchange rate regime, in: *Journal of International Money and Finance* 11, S. 340-365.
- Johnson, E. A. (1941), New tools for the economic historian, in: *Journal of Economic History (Supplement)* 1, S. 30-38.
- Johnstone, D. J. (1986), Tests of significance in theory and practice (with discussion), in: *The Statistician* 35, S. 491-504.
- Johnston, J. (1991), Econometrics retrospect and prospect, in: *Economic Journal* 101, S. 51-56. [wiederabgedruckt in Darnell (1994a)]
- John, V. (1884), *Geschichte der Statistik. Bd. 1. Von dem Ursprung der Statistik bis auf Quetelet (1835)*, Stuttgart.
- Jones, L. V. (Hg.) (1986), *The collected works of John W. Tukey. Bd. 3. Philosophy and principles of data analysis: 1949-1964. Bd. 4. Philosophy and principles of data analysis: 1965-1986*, Belmont.
- Jones, L. V. (1992), Introduction to Tukey (1962) The future of data analysis, in: Kotz, S./Johnson, N. L. (Hgg.), *Breakthroughs in statistics. Bd. 2. Methodology and distribution*, New York u. a., S. 403-407.
- Jörberg, L. (1972), *A History of Prices in Sweden, 1732 - 1914. Bd. 2. Description, Analysis*, Lund.
- Juselius, K. (1991), Comment on E. E. Leamer, 'A Bayesian perspective on inference from macroeconomic data', in: *Scandinavian Journal of Economics* 93, S. 253-258.
- Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, in: *Journal of Basic Engineering* 82, S. 35-45. [wiederabgedruckt in Harvey (1994a)]
- Kalman, R. E. (1980), A system-theoretic critique of dynamic economic models, in: *International Journal of Policy Analysis and Information Systems* 4, S. 3-22.
- Kalman, R. E. (1982a), Dynamic econometric models: a system-theoretic critique, in: Szegö, G. P. (Hg.), *New quantitative techniques for economic analysis*, New York u. a., S. 19-28.

- Kalman, R. E. (1982b), Identification from real data, in: Hazewinkel, M./Rinnooy Kan, A. H. (Hgg.), *Current developments in the interface: Economics, econometrics and mathematics*, Dordrecht, S. 161-196.
- Kalman, R. E. (1982c), Identifiability and problems of model selection in econometrics, in: Hildenbrand, W. (Hg.), *Advances in econometrics*, Cambridge.
- Kamlah, A. (1987), The decline of the Laplacian theory of probability: a study of Stumpf, von Kries, and Meinong, in: Krüger, L./Daston, L. J./Heidelberger, M. (Hgg.), *The probabilistic revolution. Bd. 1. Ideas in history*, Cambridge u. a., S. 91-116.
- Kass, R. E./Raftery, A. E. (1995), Bayes factors, in: *Journal of the American Statistical Association* 90.
- Kaufhold, K. H. (1973), Wirtschaftsgeschichte und ökonomische Theorien. Überlegungen zum Verhältnis von Wirtschaftsgeschichte und Wirtschaftstheorie am Beispiel Deutschlands, in: Schulz, G. (Hg.), *Geschichte heute. Positionen, Tendenzen und Probleme*, Göttingen, S. 256-280.
- Kellenbenz, H. (1965), Wirtschaftsgeschichte. (I) Grundlegung, in: v. Beckerath, E. et al. (Hgg.), *Handwörterbuch der Sozialwissenschaften. Bd. 12*, Stuttgart u. a., S. 124-141.
- Kellerer, H./Mahr, W./Schneider, G./Strecker, H. (Hgg.) (1963), *Oskar Anderson. Ausgewählte Schriften. 2 Bde.*, Tübingen.
- Kempthorne, O. (1971), Comment on 'Applications of statistical inference to physics', in: Godambe, V. P./Sprott, D. A. (Hgg.), *Foundations of statistical inference*, Toronto, S. 286-287.
- Kendall, M. G. (1953), The analysis of economic time-series – Part I: prices (with discussion), in: *Journal of the Royal Statistical Society A* 117, S. 11-34.
- Kendall, M. G. (1970), Review of Box and Jenkins (1970), in: *Journal of the Royal Statistical Society A* 134, S. 450-453.
- Kendall, M. G./Plackett, R. L. (Hgg.) (1977), *Studies in the history of statistics and probability. Bd. 2*, London.
- Keuzenkamp, H. A. (1995), The econometrics of the holy grail – a review of Econometrics: alchemy or science? Essays in econometric methodology, in: *Journal of Economic Surveys* 9, S. 233-248.
- Keuzenkamp, H. A./Magnus, J. R. (1995), On tests and significance in econometrics, in: *Journal of Econometrics* 67, S. 5-24.
- Keynes, J. M. (1921), *A treatise on probability*, London.
- Keynes, J. M. (1924), *A tract on monetary reform*, London.
- Keynes, J. M. (1939), Professor Tinbergen's method, in: *Economic Journal* 49, S. 558-568. [wiederabgedruckt in Darnell (1994a), Poirier (1994a) und Hendry/Morgan (1995a)]
- Keynes, J. M. (1940), Comment [zu Tinbergen (1940a)], in: *Economic Journal* 50, S. 141-156. [wiederabgedruckt in Darnell (1994a) und Poirier (1994a)]
- King, M. L. (1992), Introduction to Durbin and Watson (1950, 1951) Testing for serial correlation in least squares for regression. I, II, in: Kotz, S./Johnson, N. L. (Hgg.), *Breakthroughs in statistics. Bd. 2. Methodology and distribution*, New York u. a., S. 229-236.

- King, W. I. (1924), Principles underlying the isolation of cycles and trends, in: *Journal of the American Statistical Association* 19, S. 468-475.
- Kirchgässner, G. (1983), Ökonometrie: Datenanalyse oder Theorienüberprüfung, in: *Jahrbücher für Nationalökonomie und Statistik* 198, S. 511-538.
- Kitchin, J. (1923), Cycles and trends in economic factors, in: *Review of Economic Statistics* 5, S. 10-16.
- Klein, J. (1997), *Statistical visions of time: a history of time series analysis from 1662-1938*, Cambridge.
- Kloek, T./van Dijk, H. K. (1978), Bayesian estimates of equation systems parameters: an application of integration by Monte Carlo, in: *Econometrica* 46, S. 1-19.
- Klump, R. (1993), *Geld, Währungssystem und optimales Wachstum*, Tübingen (= Schriften zur angewandten Wirtschaftsforschung 59).
- Knapp, G. F. (1869), *Die Sterblichkeit in Sachsen nach amtlichen Quellen dargestellt*, Leipzig.
- Koch, G. G./Gillings, D. B. (1983), Inference, design based vs. model based, in: Kotz, S./Johnson, N. L. (Hgg.), *Encyclopedia of statistical sciences. Bd. 4*, New York u. a., S. 84-88.
- Kocka, J. (1972), Sozial- und Wirtschaftsgeschichte, in: Kernig, C. D. (Hg.), *Sowjet-system und Demokratische Gesellschaft. Eine vergleichende Enzyklopädie. Bd. 6*, Freiburg/Basel/Wien, S. 1-39.
- Kocka, J. (1986), *Sozialgeschichte. Begriff – Entwicklung – Probleme*, 2. Aufl., Göttingen.
- Köllman, W. et al. (1995), Wirtschafts- und Sozialgeschichte – Neue Wege? Zum wissenschaftlichen Standort des Faches, in: *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte* 82, S. 387-422.
- Komlos, J. (1995), Die Zukunft der Wirtschaftsgeschichte in Europa, in: *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte* 82, S. 404-410.
- Kondratieff, N. D. (1926), Die langen Wellen der Konjunktur, in: *Archiv für Sozialwissenschaft und Sozialpolitik* 56, S. 573-609.
- Koopman, S./Harvey, A. C./Doornik, J. A./Shephard, N. (1995), *STAMP 5.0. Structural time series analyser, modeller and predictor*, London u. a.
- Koopmans, T. C. (1937), *Linear regression analysis of economic time series*, Haarlem (= Netherlands Economic Institute, Publication No. 20). [teilweise wiederabgedruckt in Hendry/Morgan (1995a)]
- Koopmans, T. C. (1941), The logic of econometric business cycle research, in: *Journal of Political Economy* 49, S. 157-181. [wiederabgedruckt in Darnell (1994a)]
- Koopmans, T. C. (1947), Measurement without theory, in: *Review of Economics and Statistics* 29, S. 161-172. [wiederabgedruckt in Darnell (1994a), Poirier (1994a) und Hendry/Morgan (1995a)]
- Koopmans, T. C. (1949), A reply [to Vining (1949a)], in: *Review of Economics and Statistics* 31, S. 86-91. [wiederabgedruckt in Darnell (1994a), Poirier (1994a) und Hendry/Morgan (1995a)]
- Koot, G. (1987), *English historical economics, 1879-1926*, New York.
- Kotz, S./Johnson, N. L. (Hgg.) (1992a), *Breakthroughs in statistics. Bd. 1. Foundations and basic theory. Bd. 2. Methodology and distribution*, New York u. a. (= Springer Series in Statistics).

- Kotz, S./Johnson, N. L. (1992b), Preface, in: Kotz, S./Johnson, N. L. (Hgg.), *Break-throughs in statistics*. Bd. 1. Foundations and basic theory, New York u. a. (= Springer Series in Statistics), S. vii-xii.
- Krelle, W. (1959), Grundlinien einer stochastischen Konjunkturtheorie, in: *Zeitschrift für die gesamte Staatswissenschaft* 115, S. 472-491.
- Krüger, L. (1987a), The slow rise of probabilism: philosophical arguments in the nineteenth century, in: Krüger, L./Daston, L. J./Heidelberger, M. (Hgg.), *The probabilistic revolution*. Bd. 1. *Ideas in history*, Cambridge u. a., S. 59-89.
- Krüger, L. (1987b), Laplace and thereafter: The status of probability calculus in the nineteenth century, in: Krüger, L./Daston, L. J./Heidelberger, M. (Hgg.), *The probabilistic revolution*. Bd. 1. *Ideas in history*, Cambridge u. a., S. 191-214.
- Krüger, L./Daston, L. J./Heidelberger, M. (Hgg.) (1987), *The probabilistic revolution*. Bd. 1. *Ideas in history*, Cambridge u. a.
- Krüger, L./Gigerenzer, G./Morgan, M. S. (Hgg.) (1987), *The probabilistic revolution*. Bd. 2. *Ideas in the sciences*, Cambridge u. a.
- Kuhn, T. S. (1976), *Die Struktur wissenschaftlicher Revolutionen*, 2. Aufl., Frankfurt.
- Kulla, B. (1996), *Die Anfänge der empirischen Konjunkturforschung in Deutschland*, Berlin (= Volkswirtschaftliche Studien 164).
- Kusch, L. (1979), *Mathematische und naturwissenschaftliche Formeln und Tabellen*, 5. Aufl., Essen.
- Kuznets, S. S. (1928a), On moving correlation of time sequences, in: *Journal of the American Statistical Association* 23, S. 121-136.
- Kuznets, S. S. (1928b), On the analysis of time series, in: *Journal of the American Statistical Association* 23, S. 398-410.
- Kuznets, S. S. (1929), Random events and cyclical oscillations, in: *Journal of the American Statistical Association* 24, S. 258-275.
- Kuznets, S. S. (1930a), *Secular movements in production and prices*, Boston/New York.
- Kuznets, S. S. (1930b), *Wesen und Bedeutung des Trends. Zur Theorie der säkularen Bewegung*, Bonn (= Veröffentlichungen der Frankfurter Gesellschaft für Konjunkturforschung 7).
- Kuznets, S. S. (1934), Time series, in: Seligman, E. R./Johnson, A. (Hgg.), *Encyclopedia of the social sciences*. Bd. 13, New York, S. 629-636.
- Kuznets, S. S. (1940), Schumpeter's Business Cycles, in: *American Economic Review* 30, S. 257-271.
- Kuznets, S. S. (1941), Statistics and economic history, in: *Journal of Economic History (Supplement)* 1, S. 26-41.
- Kuznets, S. S. (1951), Statistical trends and historical changes, in: *Economic History Review*, 2nd ser. 3, S. 263-278.
- Kuznets, S. S. (1957), Summary of discussion and postscript to W. W. Rostow, John R. Meyer, and Alfred E. Conrad, 'The integration of economic theory and economic history', in: *Journal of Economic History* 17, S. 545-553.
- Kuznets, S. S. (1989), *Economic development, the family, and income distribution*. *Selected essays*, Cambridge u. a. (= Studies in Economic History and Policy).
- Kyburg, H. E./Smokler, H. E. (Hgg.) (1964), *Studies in subjective probability*, New York.

- Kyburg, H. (1985), Logic of statistical reasoning, in: Kotz, S./Johnson, N. L. (Hgg.), *Encyclopedia of statistical sciences*. Bd. 5, New York u. a., S. 117-122.
- Lakatos, I. (1974), Falsifikation und die Methodologie wissenschaftlicher Forschungsprogramme, in: Lakatos, I./Musgrave, A. (Hgg.), *Kritik und Erkenntnisfortschritt*, Braunschweig, S. 89-189. [Übers. von: Lakatos, I. (1970), Falsifikation and the methodology of scientific research programs, in: Lakatos, I./Musgrave, A. (Hgg.), *Criticism and the growth of knowledge*, Cambridge, S. 91-196.]
- Lamprecht, K. (1896), *Alte und neue Richtungen in der Geschichtswissenschaft*, Berlin. [teilweise wiederabgedruckt in Schleier (1988a)]
- Lamprecht, K. (1896-1897), Was ist Kulturgeschichte? Beitrag zu einer empirischen Historik, in: *Deutsche Zeitschrift für Geschichtswissenschaft*, *Neue Folge* 1, S. 75-150. [z. T. wiederabgedruckt in Schleier (1988a)]
- Lancaster, H. O. (1988), Statistics, history of, in: Kotz, S./Johnson, N. L. (Hgg.), *Encyclopedia of statistical sciences*. Bd. 8, New York u. a., S. 704-711.
- Laud, P. W./Ibrahim, J. G. (1995), Predictive model selection, in: *Journal of the Royal Statistical Society B* 57, S. 247-262.
- Lawson, T. (1989), Realism and instrumentalism in the development of econometrics, in: *Oxford Economic Papers* 41, S. 236-258.
- Lazarsfeld, P. F. (1961), Notes on the history of quantification in sociology – trends, sources and problems, in: Woolf, H. (Hg.), *Quantification – A history of the meaning of measurement in the natural and social sciences*, Indianapolis, S. 147-203.
- Leamer, E. E. (1978), *Specification searches: ad hoc inference with nonexperimental data*, New York (= Wiley Series in Probability and Mathematical Statistics).
- Leamer, E. E. (1983), Let's take the con out of econometrics, in: *American Economic Review* 73, S. 31-43. [wiederabgedruckt in Granger (1990), Poirier (1994a) und Leamer (1994a)]
- Leamer, E. E. (1985), Vector autoregressions for causal inference?, in: *Carnegie Rochester Conference Series on Public Policy* 22/255, S. 283-304. [wiederabgedruckt in Poirier (1994a)]
- Leamer, E. E. (1986), A Bayesian analysis of the determinants of inflation, in: Belsley, D. A./Kuh, E. (Hgg.), *Model reliability*, Massachusetts, S. 62-89. [wiederabgedruckt in Leamer (1994a)]
- Leamer, E. E. (1987), Econometric metaphors, in: Bewley, T. F. (Hg.), *Advances in econometrics: Fifth world congress*. Bd. 2, Cambridge, S. 1-28.
- Leamer, E. E. (1991), A Bayesian perspective of inference from macroeconomic data, in: *Scandinavian Journal of Economics* 93, S. 225-248. [wiederabgedruckt in Leamer (1994a)]
- Leamer, E. E. (Hg.) (1994a), *Sturdy econometrics*, Aldershot (= Economists of the Twentieth Century).
- Leamer, E. E. (1994b), Introduction, in: Leamer, E. E. (Hg.), *Sturdy econometrics*, Aldershot (= Economists of the Twentieth Century), S. ix-xvi.
- Leamer, E. E./Leonard, H. (1983), Reporting the fragility of regression estimates, in: *Review of Economics and Statistics* 65, S. 306-317. [wiederabgedruckt in Leamer (1994a)]

- Ledolter, J. (1989), Adaptive estimation and structural change in regression and time series models, in: Hackl, P. (Hg.), *Statistical analysis and forecasting of economic structural change*, Berlin, S. 191-208.
- Lehmann, E. L. (1964), *Testing statistical hypotheses*, 3. Aufl., New York.
- Lehmann, E. L. (1992), Introduction to Neyman and Pearson (1933) On the problem of the most efficient tests of statistical hypotheses, in: Kotz, S./Johnson, N. L. (Hgg.), *Breakthroughs in statistics. Bd. 1. Foundations and basic theory*, New York u. a, S. 67-72.
- Lehmann, E. L. (1993), The Fisher, Neyman-Pearson theories of testing hypotheses: One theory or two?, in: *Journal of the American Statistical Association* 88, S. 1242-1249.
- Leserer, M. (1986), Kognitive Inferenz als ökonometrische Aufgabe: einige Bemerkungen zur ökonometrischen Grundsatzdiskussion, in: *Jahrbücher für National-ökonomie und Statistik* 201, S. 97-106.
- Lindley, D. V. (1972), Obituary: L. J. Savage, in: *Journal of the Royal Statistical Society A* 135, S. 462-463.
- Lindley, D. V. (1980), Savage, Leonard Jimmie, in: Sills, D. L. (Hg.), *International encyclopedia of the social sciences. Biographical supplement*, New York/London, S. 691-694.
- Lindley, D. V. (1991), Statistical inference, in: Eatwell, J./Milgate, M./Newman, P. (Hgg.), *The new Palgrave. A dictionary of economics. Bd. 4*, 2. Aufl., London/New York/Tokyo, S. 490-493.
- Liu, L.-M. (1986), *Multivariate time series analysis using vector ARMA models*, De Kalb.
- Liu, L.-M. (1993), *Modeling and forecasting time series using SCA-expert capabilities*, Chicago.
- Liu, T. C. (1960), Underidentification, structural estimation, and forecasting, in: *Econometrica* 28, S. 855-865.
- Lorenz, P. (1935), Annual survey of statistical techniques, trends and seasonal variations, in: *Econometrica* 3, S. 456-471.
- Lovell, M. C. (1983), Data mining, in: *Review of Economics and Statistics* 65, S. 1-12.
- Lütkepohl, H. (1991), *Introduction to multiple time series analysis*, Berlin u. a.
- Lüthy, C. (1997), Letzte Garde der Rationalität. Der "Golem" tobt: In Amerika wird zwischen den Disziplinen ein Wissenschaftskrieg geführt, in: *Frankfurter Allgemeine Zeitung*, 9.7.
- Lütkepohl, H./Reimers, H. G. (1992), Impulse response analysis of co-integrated systems, in: *Journal of Economic Dynamics and Control* 16, S. 53-78.
- Lutz, R. R. (1949), *Graphic presentation simplified*, New York.
- Maddala, G. S. (1992), *Introduction to econometrics*, 2. Aufl., New York u. a.
- Magnus, J. R./Morgan, M. S. (1987), The ET interview: Professor J. Tinbergen, in: *Econometric Theory* 3, S. 117-142.
- Malinvaud, E. (1991), Review [von Morgan (1990)], in: *Economic Journal* 101, S. 634-636.
- Mallows, C. L./Tukey, J. W. (1982), An overview of techniques of data analysis, emphasizing its exploratory aspects, in: de Oliveira, J. T. et al. (Hgg.), *Some recent advances in statistics*, London, S. 111-172. [wiederabgedruckt in Lindley (1986)]

- Mann, H. B./Wald, A. (1943), On the statistical treatment of linear stochastic difference equations, in: *Econometrica* 11, S. 173-220.
- Marschak, J. (1950), Statistical inference in economics, in: Koopmans, T. (Hg.), *Statistical inference in dynamic economic models*, New York (= Cowles Commission Monograph 10).
- Mayr, G. (1877), *Die Gesetzmäßigkeit im Gesellschaftsleben: Statistische Studien*, München.
- McAleer, M. (1995), Sherlock Holmes and the search for truth: a diagnostic tale, in: Oxley, L./George, D. A./Roberts, C. (Hgg.), *Surveys in econometrics*, Oxford, S. 91-138.
- McCarthy, M. D. (1992), The Cowles Commission, the Brookings project, and the econometric services industry: successes and possible new directions: a personal view, in: *Econometric Theory* 8, S. 383-401.
- McClelland, P. D. (1975), *Causal explanation and model building in history, economics, and new economic history*, Ithaca, N. Y.
- McCloskey, D. N. (1983), The rhetoric of economics, in: *Journal of Economic Literature* 21, S. 481-517.
- McCloskey, D. N. (1985a), *The rhetoric of economics*, Madison.
- McCloskey, D. N. (1985b), The loss function has been mislaid: the rhetoric of significance tests, in: *American Economic Review. Papers and Proceedings* 75, S. 201-205.
- McCloskey, D. N. (1987), *Econometric history*, London.
- McCloskey, D. N. (1990), *If you're so smart: the narrative of economic expertise*, Chicago.
- McCloskey, D. N. (1993), *Knowledge and persuasion in economics*, Cambridge.
- Mellor, D. H. (1982), Chance (II), in: Kotz, S./Johnson, N. L. (Hgg.), *Encyclopedia of statistical sciences. Bd. I*, New York u. a., S. 405-411.
- Ménard, C. (1987), Why was there no probabilistic revolution in economic thought?, in: Krüger, L./Gigerenzer, G./Morgan, M. S. (Hgg.), *The probabilistic revolution. Bd. 2. Ideas in the sciences*, Cambridge u. a., S. 139-146. [wiederabgedruckt in Darnell (1994a)]
- Menges, G. (1959a), *Stichproben aus endlichen Gesamtheiten*, Frankfurt am Main (= Frankfurter Wissenschaftliche Beiträge, Rechts- und Wirtschaftswissenschaftliche Reihe 17).
- Menges, G. (1959b), Zur stochastischen Grundlegung der Ökonometrie, in: *Zeitschrift für die gesamte Staatswissenschaft* 115, S. 611-625.
- Menges, G. (1961), *Ökonometrie*, Wiesbaden.
- Menges, G. (1965), Über Wahrscheinlichkeitsinterpretationen, in: *Statistische Hefte* 6, S. 81-96.
- Menges, G. (1967a), Die Überwindung der Ungewißheit, in: N. N. (Hg.), *Wissenschaft und Praxis. Festschrift zum zwanzigjährigen Bestehen des Westdeutschen Verlages*, Köln/Opladen, S. 357-387.
- Menges, G. (1967b), Über Thomas Bayes (1702-1761) und das Theorem – Versuch einer Würdigung, in: Diemer, A. (Hg.), *Geschichte und Zukunft. Anton Hain zum 75. Geburtstag*, Meisenheim/Glan, S. 485-498.
- Menges, G. (1972), *Grundriß der Statistik. Teil 1: Theorie*, 2. Aufl., Opladen.

- Menges, G. (1976), Deskription und Inferenz (Moderne Aspekte der Frankfurter Schule), in: *Allgemeines Statistisches Archiv* 60, S. 290-319.
- Menges, G. (1980a), Ätialität und Adäquation. Dem Andenken an Heinrich Hartwich (1907-1981), in: *Statistische Hefte* 22, S. 144-150.
- Menges, G. (1980b), Adaptive Statistik. Bemerkungen über neuere Bestrebungen in der statistischen Methodologie, in: *Statistische Hefte* 21, S. 182-208.
- Menges, G./Skala, H. J. (1973), *Grundriß der Statistik. Teil 2: Daten. Ihre Gewinnung und Verarbeitung*, Opladen.
- Meschkowski, H./Schütz, H. (1972), *Schülerduden. Die Mathematik I (Bis 10. Schuljahr)*, 3. Aufl., Mannheim/Wien/Zürich.
- Metz, R. (1983), 'Long waves' in English and German economic historical series from the middle of the sixteenth to the middle of the twentieth century, in: Fremdling, R./O'Brian, P. (Hgg.), *Productivity in the economies of Europe*, Stuttgart (= Historisch-Sozialwissenschaftliche Forschungen 15), S. 175-219.
- Metz, R. (1984), Zur empirischen Evidenz "langer Wellen", in: *Kyklos* 37, S. 266-290.
- Metz, R. (1987), Kondratieff and the theory of linear filters, in: Vasko, T. (Hg.), *The long-wave debate*, Berlin u. a., S. 390-405.
- Metz, R. (1988a), Erkenntnisziele zeitreihenanalytischer Forschung, in: *Historical Social Research/Historische Sozialforschung* 13/3, S. 6-22.
- Metz, R. (1988b), Ansätze, Begriffe und Verfahren der Analyse ökonomischer Zeitreihen, in: *Historical Social Research/Historische Sozialforschung* 13/3, S. 23-103.
- Metz, R. (1992a), A re-examination of long waves in aggregate production series, in: Kleinknecht, A./Mandel, E./Wallerstein, I. (Hgg.), *New findings in long wave research*, London, S. 80-119.
- Metz, R. (1992b), Über die stochastische Struktur langfristiger Wachstumsschwankungen, in: *IFO-Studien* 38, S. 171-197.
- Metz, R. (1993), Probleme der statistischen Analyse langer historischer Zeitreihen, in: *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte* 80, S. 457-486.
- Metz, R. (1995), *Stochastische Trends und langfristige Wachstumsschwankungen. Neue Forschungsansätze und ihre theoretische und empirische Relevanz für die Wirtschaftsgeschichte*, Habil. St. Gallen.
- Metz, R./Spree, R. (1981), Kuznets-Zyklen im Wachstum der deutschen Wirtschaft während des 19. und frühen 20. Jahrhunderts, in: Petzina, D./van Roon, G. (Hgg.), *Konjunktur, Krise, Gesellschaft. Wirtschaftliche Wechsellagen und soziale Entwicklung im 19. und 20. Jahrhundert*, Stuttgart (= Bochumer Historische Studien 25), S. 343-378.
- Metz, R./Stier, W. (1992a), Modelling the long wave-phenomena, in: *Historical Social Research/Historische Sozialforschung* 17, S. 43-62.
- Metz, R./Stier, W. (1992b), Filter-design in the frequency domain, in: Kleinknecht, A./Mandel, E./Wallerstein, I. (Hgg.), *New findings in long wave research*, London, S. 45-79.
- Metz, R./Thome, H. (1991), Verläufe der langfristige Wertewandel zyklisch? Eine zeitreihenanalytische Überprüfung der Thesen von Namenwirth und Weber, in: Best, H./Thome, H. (Hgg.), *Neue Methoden der Analyse historischer Daten*, St. Katharinen (= Historisch Sozialwissenschaftliche Forschungen 23), S. 269-301.

- Mills, T. C. (1992), *Predicting the unpredictable? Science and guesswork in financial market forecasting*, London (= Occasional Paper/IEA 87).
- Mills, T. C. (1993), *The econometric modelling of financial time series*, Cambridge.
- Mirowski, P. (1981), The rise (and retreat) of a market: English joint stock shares in the eighteenth century, in: *Journal of Economic History* 41, S. 559-577.
- Mirowski, P. (1987), What do markets do? Efficiency tests of the 18th-century London stock market, in: *Explorations in Economic History* 24, S. 107-129.
- Mirowski, P. (1989a), *More heat than light: economics as social physics; physics as nature's economics*, Cambridge (= Historical Perspectives on Modern Economics).
- Mirowski, P. (1989b), The measurement without theory controversy: defeating rival research programs by accusing them of the naive empiricism, in: *Economies et Sociétés (Cahiers de l'ISMEA)* 23, S. 65-87.
- Mirowski, P. (1989c), The probabilistic counter-revolution, or how stochastic concepts came to neoclassical economic theory, in: *Oxford Economic Papers* 41, S. 217-235.
- Mirowski, P. (1990), From Mandelbrot to chaos in economic theory, in: *Southern Economic Journal* 57, S. 289-307.
- Mirowski, P. (1991), The when, the how and the why of mathematical expression in the history of economic analysis, in: *The Journal of Economic Perspectives* 5, S. 145-157.
- Mirowski, P. (Hg.) (1994), *Edgeworth on chance, economic hazard, and statistics*, Lanham/London (= The Worldly Philosophy: Studies at the Intersection of Philosophy and Economics).
- Mitchell, W. C. (1913), *Business cycles and their causes*, Berkeley (= California University Memoirs 3). [teilweise wiederabgedruckt in Hendry/Morgan (1995a)]
- Mitchell, W. C. (1925), Quantitative analysis in economic theory, in: *American Economic Review* 15, S. 1-12. [wiederabgedruckt in Darnell (1994a)]
- Mitchell, W. C. (1927), *Business cycles: the problem and its setting*, New York (= National Bureau of Economic Research 10). [deutsche, vom Verfasser durchgesehene und ergänzte Übers. der 2. Aufl. Leipzig 1931]
- Möller, J./Wais, B. (1987), Kalman-Verfahren in der Ökonometrie. Schätzung einer Geldangebotsgleichung mit zeitvariablen Koeffizienten unter Verwendung optimaler Filtereigenschaften, in: *Allgemeines Statistisches Archiv* 71, S. 267-283.
- Monz, J. (Hg.) (1991), *Econometrics: Essays on theory and applications. Collected papers of Franklin M. Fisher*, New York u. a.
- Moore, G. H./Zarnowitz, V. (1986), The development and role of the National Bureau of Economic Research's Business cycle chronologies, in: Gordon, R. J. (Hg.), *The American business cycle: Continuity and change*, Chicago/London (= National Bureau of Economic Research Studies in Business Cycles 25), S. 735-779.
- Moore, H. L. (1911), *Laws of wages: an essay in statistical economics*, New York.
- Moore, H. L. (1914), *Economic cycles – their law and cause*, New York. [teilweise wiederabgedruckt in Hendry/Morgan (1995a)]
- Moore, H. L. (1923), *Generating economic cycles*, New York.
- Mooser, J. (1990), Wirtschafts- und Sozialgeschichte, Historische Sozialwissenschaft, Gesellschaftsgeschichte, in: van Dülmen, R. (Hg.), *Fischer Lexikon Geschichte*, Frankfurt a. M., S. 86-101.

- Morgan, M. S. (1987a), The probabilistic revolution in economics – an overview, in: Krüger, L./Gigerenzer, G./Morgan, M. S. (Hgg.), *The probabilistic revolution. Bd. 2. Ideas in the sciences*, Cambridge u. a., S. 135-137.
- Morgan, M. S. (1987b), Statistics without probability and Haavelmo's revolution in econometrics, in: Krüger, L./Gigerenzer, G./Morgan, M. S. (Hgg.), *The probabilistic revolution. Bd. 2. Ideas in the sciences*, Cambridge u. a., S. 171-197. [wiederabgedruckt in Darnell (1994a)]
- Morgan, M. S. (1988), Finding a satisfactory empirical model, in: de Marchi, N. (Hg.), *The Popperian legacy in economics. Papers presented at a symposium in Amsterdam, December 1985*, Cambridge u. a., S. 199-211.
- Morgan, M. S. (1990), *The history of econometric ideas*, Cambridge u. a.
- Morgan, M. S. (1991), The stamping out of process analysis in econometrics, in: de Marchi, N./Blaug, M. (Hgg.), *Appraising economic theories: studies in the methodology of research programs*, Aldershot, S. 237-265.
- Morgan, M. S. (1995a), The history of econometrics: errors and refutations, in: *Econometric Theory* 11, S. 392-397.
- Morgan, M. S. (1995b), Empirical model particularities and belief in the natural rate hypothesis, in: *Journal of Econometrics* 67, S. 81-102.
- Morgenstern, O. (1952), *Über die Genauigkeit wirtschaftlicher Beobachtungen*, München (= Einzelschriften der Deutschen Statistischen Gesellschaft 4).
- Morris, C. (1983), Parametric empirical Bayes inference: theory and applications (with discussion), in: *Journal of the American Statistical Association* 78, S. 47-65.
- Mosteller, F./Tukey, J. W. (1968), Data analysis, including statistics, in: Lindzey, G./Aronson, E. (Hgg.), *The handbook of social psychology. Bd. 2. Research methods*, 2. Aufl., Reading, Mass. u. a., S. 80-203. [teilweise wiederabgedruckt in Jones (1986) und in Cleveland (1988a)]
- Mosteller, F./Wallace, D. L. (1964), *Inference and disputed authorship: The Federalist*, Reading, Mass.
- N. N. (1985), Models I, II, and III, in: Kotz, S./Johnson, N. L. (Hgg.), *Encyclopedia of statistical sciences. Bd. 5*, New York u. a., S. 589-590.
- Nardinelli, C./Meiners, R. E. (1988), Schmoller, the Methodenstreit, and the development of economic history, in: *Journal of Institutional and Theoretical Economics (Views and comments on Gustav Schmoller and the Methodenstreit)* 144, S. 543-551.
- Neal, L. (1985), Integration of international capital markets: quantitative evidence from the eighteenth to twentieth centuries, in: *Journal of Economic History* 45, S. 219-226.
- Neal, L. (1987), The integration and efficiency of the London and Amsterdam stock markets in the eighteenth century, in: *Journal of Economic History* 47, S. 97-115.
- Neal, L. (1990), *The rise of financial capitalism: international capital markets in the age of reason*, New York.
- Neal, L. (1991), The Dutch and English East India companies compared: evidence from the stock and foreign exchange markets, in: Tracy, J. D. (Hg.), *The Rise of merchant empires: long distance trade in the early modern world, 1350 -1750*, 2. Aufl., Cambridge u. a., S. 195-223.

- Neal, L. (1994), The finance of business during the industrial revolution, in: Floud, R./McCloskey, D. (Hgg.), *The economic history of Britain since 1700. Bd. 1. 1700-1860*, 2. Aufl., Cambridge, S. 151-181.
- Nelson, C. R. (1972), The prediction performance of the FRB-MIT-PENN model of the U.S. economy, in: *American Economic Review* 62, S. 902-917. [wiederabgedruckt in Wallis (1994a)]
- Nelson, C. R./Plosser, C. I. (1982), Trends and random walks in macroeconomic time series, in: *Journal of Monetary Economics* 10, S. 139-162.
- Nerlove, M. (1966), A tabular survey of macro-econometric models, in: *International Economic Review* 7, S. 127-175. [wiederabgedruckt in Wallis (1994a)]
- Nerlove, M./Grether, D. M./Carvalho, J. L. (1979), *Analysis of economic time series: a synthesis*, New York.
- Neuberg, L. G. (1995), Book review: the history of econometric ideas by Mary S. Morgan, in: *Econometric Theory* 11, S. 371-383.
- Newton, H. J. (1993), Graphics for time series analysis, in: Rao, C. R. (Hg.), *Handbook of statistics. Bd. 9*, New York, S. 803-823.
- Neyman, J. (1971), Foundations of behavioristic statistics, in: Godambe, V. P./Sprott, D. A. (Hgg.), *Foundations of statistical inference*, Toronto, S. 1-13.
- Neyman, J. (1981), Introduction, in: Ondar, K. (Hg.), *Correspondence between A. A. Markow and A. A. Chuprov on the theory of probability and mathematical statistics*, New York, S. v-viii.
- Nocken, U. (1993), Die Große Deflation: Goldstandard, Geldmenge und Preise in den USA und Deutschland 1870 bis 1896, in: Schremmer, E. (Hg.), *Geld und Währung vom 16. Jahrhundert bis zur Gegenwart*, Stuttgart (= VSWG Beihefte 106), S. 157-189.
- North, D. C. (1968), History: VII. Economic history, in: Sills, D. L. (Hg.), *International encyclopedia of the social sciences. Bd. 6*, New York, S. 468-474.
- O'Hagan, A. (1995), Fractional Bayes factors for model comparison (with discussion), in: *Journal of the Royal Statistical Society B* 57, S. 99-138.
- Odell, K. A. (1989), The integration of regional and interregional capital markets: evidence from the Pacific Coast, 1883-1913, in: *Journal of Economic History* 49, S. 297-310.
- Ogburn, W. F./Thomas, D. S. (1922), The influence of the business cycle on certain social conditions, in: *Journal of the American Statistical Association* 18, S. 324-340.
- Ohler, N. (1980), *Quantitative Methoden für Historiker. Eine Einführung. Mit einer Einführung in die EDV von Hermann Schäfer*, München.
- Oxley, L./George, D. A./Roberts, C. J. (Hgg.) (1995), *Surveys in econometrics*, Oxford.
- Pagan, A. (1987), Three econometric methodologies: a critical appraisal, in: *Journal of Economic Surveys* 1, S. 3-24. [wiederabgedruckt in Granger (1990), Oxley/George/Roberts (1995) und Poirier (1994a)]
- Pagan, A. (1995), Three econometric methodologies: an update, in: Oxley, L./George, D. A./Roberts, C. J. (Hgg.), *Surveys in econometrics*, Oxford, S. 30-41.

- Parenti, G. (1942), *Prezzi e mercato del grano a Siena (1564-1765)*, Florenz (= Pubblicazioni della R. Università degli Studi di Firenze. Facoltà di Economia e Commerciale 19. Scuola di Statistica). [Neudruck Paris 1981]
- Pearson, E. S. (1966), The Neyman-Pearson story: 1926-34, in: David, F. N. (Hg.), *Research papers in statistical: Festschrift for J. Neyman*, New York. [wiederabgedruckt in Pearson/Kendall (1970)]
- Pearson, E. S. (1967), Some reflections on continuity in the development of mathematical statistics, 1885-1920, in: *Biometrika* 52, S. 3-18. [wiederabgedruckt in Pearson/Kendall (1970)]
- Pearson, E. S./Kendall, M. G. (Hgg.) (1970), *Studies in the history of statistics and probability*, London.
- Pearson, K. (1898), Mathematical contributions to the theory of evolution: on the law of ancestral heredity, in: *Proceedings of the Royal Society of London* 62, S. 386-412.
- Pearson, K./Filon, L. N. (1898), Mathematical contributions to the theory of evolution, IV: on the probable errors of frequency constants and on the influence of random selection on variation and correlation, in: *Philosophical Transactions of the Royal Society of London (A)* 191, S. 229-311.
- Pelloni, G. (1987), A note on Friedman and the Neo-Bayesian approach, in: *Manchester School of Economics and Social Studies* 55, S. 407-418.
- Persons, W. M. (1910), The correlation of economic statistics, in: *Journal of the American Statistical Association* 12, S. 287-322. [teilweise wiederabgedruckt in Hendry/Morgan (1995a)]
- Persons, W. M. (1919a), Indices of business conditions, in: *Review of Economic Statistics* 1, S. 5-110.
- Persons, W. M. (1919b), An index of general business conditions, in: *Review of Economic Statistics* 1, S. 111-205.
- Persons, W. M. (1924a), Some fundamental concepts of statistics, in: *Journal of the American Statistical Association* 19, S. 1-8.
- Persons, W. M. (1924b), The problem of business forecasting, in: N. N. (Hg.), *The problem of business forecasting*, London (= Pollak Foundation for Economic Research Publications 6).
- Persons, W. M. (1925), Statistics and economic theory, in: *Review of Economic Statistics* 7, S. 179-197. [wiederabgedruckt in Darnell (1994a)]
- Pesaran, M. H. (1991), Econometrics, in: Eatwell, J./Milgate, M./Newman, P. (Hgg.), *The new Palgrave. A dictionary of economics. Bd. 2, 2. Aufl.*, London/New York/Tokyo, S. 8-22.
- Pesaran, M. H./Pesaran, B. (1991), *MICROFIT 3.0. An interactive econometric software package. User manual*, Oxford.
- Pesaran, M. H./Smith, R. P. (1985), Keynes on econometrics, in: Lawson, T./Pesaran, M. H. (Hgg.), *Keynes economics: methodological issues*, London, S. 134-150.
- Phillips, P. C. (1986), Understanding spurious regression in econometrics, in: *Journal of Econometrics* 33, S. 311-340.
- Phillips, P. C. (1995a), Bayesian model selection and prediction with empirical applications, in: *Journal of Econometrics* 69, S. 289-331.
- Phillips, P. C. (1995b), Bayesian prediction. A response, in: *Journal of Econometrics* 69, S. 351-365.

- Phillips, P. C. (1996), Econometric model determination, in: *Econometrica* 64, S. 763-812.
- Poirier, D. J. (Hg.) (1994a), *The methodology of econometrics. 2 Bde.*, Aldershot (= The International Library of Critical Writings in Econometrics 6).
- Poirier, D. J. (1994b), Introduction, in: Poirier, D. J. (Hg.), *The methodology of econometrics. Bd. 1*, Aldershot (= The International Library of Critical Writings in Econometrics 6), S. xi-xxii.
- Pole, A./West, M./Harrison, J. (1994), *Applied Bayesian forecasting and time series analysis*, New York.
- Polson, N. G./Tiao, G. C. (Hgg.) (1995a), *Bayesian inference. 2 Bde.*, Aldershot (= The International Library of Critical Writings in Econometrics 7).
- Polson, N. G./Tiao, G. C. (1995b), Introduction, in: Polson, N. G./Tiao, G. C. (Hgg.), *Bayesian inference. Bd. 1*, Aldershot (= The International Library of Critical Writings in Econometrics 7), S. ix-xv.
- Popper, K. R. (1993), *Objektive Erkenntnis. Ein evolutionärer Entwurf*, Hamburg.
- Popper, K. R. (1994), *Logik der Forschung*, 10. Aufl., Tübingen (= Die Einheit der Gesellschaftswissenschaften 4).
- Popper, K. R. (1995), Propensitäten, Wahrscheinlichkeiten und die Quantentheorie (1957), in: Miller, D. (Hg.), *Karl R. Popper Lesebuch. Ausgewählte Texte zu Erkenntnistheorie, Philosophie der Naturwissenschaften, Metaphysik, Sozialphilosophie*, Tübingen.
- Popper, K. R. (1996), *Eine Welt der Propensitäten*, Tübingen.
- Porter, T. M. (1986), *The rise of statistical thinking, 1820-1900*, Princeton.
- Porter, T. M. (1987), Lawless society: social science and the reinterpretation of statistics in Germany, 1850-1880, in: Krüger, L./Daston, L. J./Heidelberger, M. (Hgg.), *The probabilistic revolution. Bd. 1. Ideas in history*, Cambridge u. a., S. 351-375.
- Porter, T. M. (1993), Interpreting the triumph of mathematical economics, in: de Marchi, N. (Hg.), *Non-natural social sciences: Reflecting on the enterprise of 'More heat than light'*, Durham/London (= Supplement to History of Political Economy 25), S. 54-68.
- Postan, M. M. (1957), N. S. B. Gras – A memoir, in: *Economic History Review, 2nd ser.* 9, S. 485.
- Pötscher, B. M. (1991), Effects of model selection on inference, in: *Econometric Theory* 7, S. 163-185.
- Poynting, P. (1884), On the fluctuation of wheat prices, in: *Journal of the Royal Statistical Society*.
- Pratt, J. W. (1965), Bayesian interpretation of standard inference statements (with discussion), in: *Journal of the Royal Statistical Society B* 27, S. 169-203.
- Pratt, J. W. (1971), Comment on: 'Probability, statistics and knowledge business' by O. Kempthorne, in: Godambe, V. P./Sprott, D. A. (Hgg.), *Foundations of statistical inference*, Toronto.
- Qin, D. (1993), *The formation of econometrics: a historical perspective*, Oxford.
- Quenouille, M. H. (1957), *The analysis of multiple time-series*, London.
- Raftery, A. E. (1995), Bayesian model selection in social research (with discussion), in: Marsden, P. V. (Hg.), *Sociological methodology 1995*, New York.

- Rahlf, T. (1993a), Ein einfaches Programm zur Berechnung von gewogenen gleitenden Mittelwerten in dBASE, in: *Historical Social Research/Historische Sozialforschung* 18/1, S. 122-144.
- Rahlf, T. (1993b), Ausgewählte Methoden der konjunkturstatistischen Untersuchung am Beispiel des Lübecker Häusermarktes von 1284 bis 1700, in: Hammel-Kiesow, R. (Hg.), *Wege zur Erforschung städtischer Häuser und Höfe. Beiträge zur fächerübergreifenden Zusammenarbeit am Beispiel Lübecks im Spätmittelalter und in der frühen Neuzeit*, Neumünster, S. 374-380.
- Rahlf, T. (1994), PC-Programme zur Zeitreihenanalyse: Datenmanagement, Grafik und univariate Analyseverfahren (SPSS, SYSTAT, STATISTICA, MicroTSP, Mesosaur), in: *Historical Social Research/Historische Sozialforschung* 19/3, S. 78-123.
- Rahlf, T. (1995a), Neue Literatur zur statistischen Graphik und graphisch gestützten Datenanalyse, in: *ZUMA-Nachrichten* 36, S. 151-165.
- Rahlf, T. (1995b), Statistical software computing reviews: MicroFit 3.21, in: *Historical Social Research/Historische Sozialforschung* 20/3, S. 88-101.
- Rahlf, T. (1996a), Durchbrüche in der modernen Statistik, in: *ZUMA-Nachrichten* 38, S. 138-147.
- Rahlf, T. (1996b), *Zur Modellierung stochastischer Zyklen als AR(2)-Prozeß*, Halle (= Der Hallesche Graureiher 96-3. Forschungsberichte des Instituts für Soziologie).
- Rahlf, T. (1996c), *The use of time in time series analysis of historical processes: some critical remarks to Isaac and Griffin*, o. O. (= Paper presented at the 1st European Social Science History Conference, 11. May 1996, Leeuwenhorst, Netherlands).
- Rahlf, T. (1996d), *Getreide in der Sozial- und Wirtschaftsgeschichte vom 16. bis 18. Jahrhundert. Das Beispiel Köln im regionalen Vergleich*, Trier (= Kleine Schriften zur Geschichte und Landeskunde 3).
- Ramsey, F. P. (1931a), Truth and probability (1926), in: Braithwaite, R. B. (Hg.), *The foundations of mathematics and other logical essays by Frank Plumpton Ramsey*, London (= International Library of Psychology, Philosophy and Scientific Method). [wiederabgedruckt in Kyburg/Smokler (1964) und Polson/Tiao (1995a)]
- Ramsey, F. P. (1931b), Further considerations (1928), in: Braithwaite, R. B. (Hg.), *The foundations of mathematics and other logical essays by Frank Plumpton Ramsey*, London (= International Library of Psychology, Philosophy and Scientific Method), S. 199-211.
- Redlich, F. (1965), New and traditional approaches to economic history, in: *Journal of Economic History* 25, S. 480-495. [dt. Übers. in Wehler (1973a)]
- Regan, F. (1936), The admissibility of time series, in: *Econometrica* 4, S. 189.
- Reinhardt, F./Soeder, H. (1977), *dtv-Atlas zur Mathematik. Bd. 2. Analysis und angewandte Mathematik*, München.
- Reiß, W. (1985), Historische Wechselkurse, in: *Bankhistorisches Archiv*, S. 3-40.
- Reiß, W. (1986), Historical exchange rates, in: Fischer, W./McInnis, R. M./Schneider, J. (Hgg.), *The Emergence of a World Economy 1500-1914. Papers of the IX. International Congress of Economic History. Part I: 1500-1850*, Wiesbaden (= Beiträge zur Wirtschafts- und Sozialgeschichte 33, I), S. 171-190.
- Rider, C. (1995), *An introduction to economic history*, Cincinnati.

- Robbins, H. (1955), An empirical Bayes approach to statistics, in: *Proceedings of the 3rd Berkeley Symposium on Mathematical and Statistical Probability* 1, S. 157-163. [wiederabgedruckt in Kotz/Johnson (1992a)]
- Robbins, L. (1932), *An essay on the nature and significance of economic science*, London. [teilweise wiederabgedruckt in Hendry/Morgan (1995a)]
- Rogers, J. E. (1866-1902), *A history of agriculture and prices. 8 Bde.*, London.
- Rommelfanger, R. (1986), *Differenzengleichungen*, Mannheim/Wien/Zürich.
- Roos, C. F. (1936), Annual survey of statistical techniques: the correlation and analysis of time series, part II, in: *Econometrica* 4, S. 368-381.
- Rossi, P. (1986), Max Weber und die Methodologie der Geschichts- und Sozialwissenschaften, in: Kocka, J. (Hg.), *Max Weber, der Historiker*, Göttingen, S. 28-50.
- Rothenberg, T. J. (1975), Bayesian analysis of simultaneous equation models, in: Fienberg, S. E./Zellner, A. (Hgg.), *Studies in Bayesian econometrics and statistics in honor of Leonard J. Savage*, Amsterdam (= Contributions to Economic Analysis 86), S. 405-424.
- Rothenberg, W. B. (1985), The emergence of a capital market in rural Massachusetts, 1730-1838, in: *Journal of Economic History* 45, S. 781-808.
- Rutsch, M. (1985), Exploration and inference, in: Schneeweiß, H./Strecker, H. (Hgg.), *Contributions to econometrics and statistics today*, Berlin u. a., S. 192-206.
- Salin, E. (1921), Zur Methode und Aufgabe der Wirtschaftsgeschichte, in: *Schmollers Jahrbuch* 45, S. 483-505.
- Sarrazin, T. (1974), *Ökonomie und Logik der historischen Erklärung. Zur Wissenschaftslogik der New Economic History*, Bonn-Bad Godesberg (= Schriftenreihe des Forschungsinstituts der Friedrich-Ebert-Stiftung 109).
- Sasuly, M. (1936), A method of smoothing economic series by moving averages, in: *Econometrica* 4, S. 206.
- Savage, L. J. (1954), *The foundations of statistics*, New York.
- Savage, L. J. (1976), On rereading R. A. Fisher (with discussion), in: *Annals of Statistics* 4, S. 441-500.
- Savage, L. J. et al. (1962), *The foundations of statistical inference. A discussion*, London.
- Scherff, H. (1980), Wahrscheinlichkeitstheorie, in: Albers, W. et al. (Hgg.), *Handwörterbuch der Wirtschaftswissenschaft. Bd. 8*, Stuttgart.
- Schleier, H. (Hg.) (1988a), *Karl Lamprecht: Alternative zu Ranke. Schriften zur Geschichtstheorie*, Leipzig.
- Schleier, H. (1988b), Der Kulturhistoriker Karl Lamprecht, der 'Methodenstreit' und die Folgen, in: Schleier, H. (Hg.), *Karl Lamprecht: Alternative zu Ranke. Schriften zur Geschichtstheorie*, Leipzig, S. 7-45.
- Schlittgen, R. (1996), *Statistische Inferenz*, München/Wien.
- Schlittgen, R./Streitberg, B. H. (1994), *Zeitreihenanalyse*, 5. Aufl., München/Wien.
- Schmitt, G. (1967), Johann Friedrich Unger (1714-1781). Arithmetiker und erster bedeutender landwirtschaftlicher Marktforscher deutscher Sprache, in: *Agrarwirtschaft. Zeitschrift für Betriebswirtschaft und Marktforschung* 16, S. 201-206.
- Schmoller, G. (1911), Volkswirtschaft, Volkswirtschaftslehre und -methode, in: Conrad, J. et al. (Hgg.), *Handwörterbuch der Staatswissenschaften. Bd. 8, 3. Aufl.*, Jena, S. 426-501.

- Schneider, I. (1987), Laplace and thereafter: the status of probability calculus in the nineteenth century, in: Krüger, L./Daston, L. J./Heidelberger, M. (Hgg.), *The probabilistic revolution. Bd. 1. Ideas in history*, Cambridge u. a., S. 191-214.
- Schneider, I. (Hg.) (1988), *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933. Einführungen und Texte*, Darmstadt.
- Schneider, W. (1986), *Der Kalmanfilter als Instrument zur Diagnose und Schätzung variabler Parameter in ökonometrischen Modellen*, Heidelberg/Wien (= Arbeiten zur Angewandten Statistik 27).
- Schnell, R. (1994), *Graphisch gestützte Datenanalyse*, München/Wien.
- Schnell, R./Hill, P. B./Esser, E. (1993), *Methoden der empirischen Sozialforschung*, 4. Aufl., München/Wien.
- Scholtz, G. (1991), *Zwischen Wissenschaftsanspruch und Orientierungsbedürfnis. Zu Grundlage und Wandel der Geisteswissenschaften*, Frankfurt a. M.
- Schulte, H. (1981), Ein neuer statistischer Ansatz zur Identifizierung von Wellenbewegungen in der langfristigen Wirtschaftsentwicklung, in: Petzina, D./v. Roon, G. (Hgg.), *Konjunktur, Krise, Gesellschaft. Wirtschaftliche Wechsellagen und soziale Entwicklungen im 19. und 20. Jahrhundert*, Stuttgart, S. 300-322.
- Schultz, H. (1928), *Statistical laws of demand and supply with special application to sugar*, Chicago.
- Schultz, H. (1934), Discussion of the question 'Is the theory of harmonic oscillations useful in the study of business cycles?', in: *Econometrica* 2, S. 189.
- Schumpeter, J. A. (1926), Gustav von Schmoller und die Probleme von heute, in: *Schmollers Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft* 50, S. 337-388.
- Schumpeter, J. A. (1933), The common sense of econometrics, in: *Econometrica* 1, S. 5-12. [wiederabgedruckt in Hendry/Morgan (1995a)]
- Schumpeter, J. A. (1939), *Business cycles: a theoretical, historical and statistical analysis of the capitalist process. 2 Bde.*, New York. [deutsche Übers.: Konjunkturzyklen. Eine theoretische, historische und statistische Analyse des kapitalistischen Prozesses, Göttingen 1961 (= Grundriß der Sozialwissenschaft)]
- Schumpeter, J. A. (1965), *Geschichte der ökonomischen Analyse*, Göttingen.
- Schuster, A. (1897), On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena, in: *Terrestrial Magnetism* 3, S. 13-41.
- Seidenfeld, T. (1979), *Philosophical problems of statistical inference: learning from R. A. Fisher*, London.
- Simon, H. A. (1957), *Models of man*, New York.
- Sims, C. A. (1972), Money, income and causality, in: *American Economic Review* 62, S. 540-552. [wiederabgedruckt in Harvey (1994a)]
- Sims, C. A. (1980a), Macroeconomics and reality, in: *Econometrica* 48, S. 1-48. [wiederabgedruckt in Granger (1990) und Poirier (1994a)]
- Sims, C. A. (1980b), Comparison of interwar and postwar business cycles: monetarism reconsidered, in: *American Economic Review* 70, S. 250-257.
- Sims, C. A. (1982), Policy analysis with econometric models (with comments), in: *Brooking Papers in Economic Activity* 1, S. 107-164.

- Sims, C. A. (1996), Macroeconomics and methodology, in: *Journal of Economic Perspectives* 10, S. 105-120.
- Sims, C. A./Uhlig, H. (1991), Understanding unit rooters: a helicopter tour, in: *Econometrica* 59, S. 1591-1599.
- Slutzky, E. (1937), The summation of random causes as the source of cyclic processes, in: *Econometrica* 5, S. 105-146.
- Snow, C. P. (1959), *The two cultures and the scientific revolution*, Cambridge.
- Sokal, A. (1996), Transgressing the boundaries: Toward a transformative hermeneutics of quantum gravity, in: *Social Text* 46/47, S. 217-252.
- Sombart, W. (1929), Economic theory and economic history, in: *Economic History Review*, 2nd ser. 2, S. 1-19.
- Sowell, F. (1991), On DeJong and Whiteman's Bayesian inference for the unit root model, in: *Journal of Monetary Economics* 28, S. 255-263.
- Spanos, A. (1986), *Statistical foundations of econometric modelling*, Cambridge.
- Spree, R. (1977), *Die Wachstumszyklen der deutschen Wirtschaft von 1840 bis 1880*, Berlin.
- Spree, R. (1978), *Wachstumstrends und Konjunkturzyklen in der deutschen Wirtschaft von 1820 bis 1913*, Göttingen.
- Spree, R. (1991), *Lange Wellen wirtschaftlicher Entwicklung in der Neuzeit. Historische Befunde, Erklärungen und Untersuchungsmethoden*, Köln (= Historical Social Research/Historische Sozialforschung Supplement 4).
- Sprott, D. A. (1965), Statistical estimation – some approaches and controversies, in: *Statistische Hefte* 6, S. 97-111.
- Stegmüller, W. (1971), Das Problem der Induktion: Humes Herausforderung und moderne Antworten, in: Lenk, H. (Hg.), *Neue Aspekte der Wissenschaftstheorie*, Braunschweig, S. 13-74.
- Stegmüller, W. (1973), *Personelle und Statistische Wahrscheinlichkeit. Erster Halbband: Personelle Wahrscheinlichkeit und Rationale Entscheidung. Zweiter Halbband. Statistisches Schließen, Statistische Begründung, Statistische Analyse*, Berlin/Heidelberg/New York (= Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie IV).
- Stephan, F. F. (1948), History of the uses of modern sampling procedures, in: *Journal of the American Statistical Association* 43, S. 12-39.
- Stigler, G. J. (1954), The early history of empirical studies of consumer behavior, in: *Journal of Political Economy* 62, S. 95-113.
- Stigler, S. (1986), *The history of statistics: the measurement of uncertainty before 1900*, Cambridge, Mass.
- Stone, R. (1978), Keynes, political arithmetic and econometrics, in: *Proceedings of the British Academy* 64, S. 55-92. [wiederabgedruckt in Darnell (1994a)]
- Stone, R. (1988), Some seventeenth century econometrics: consumer's behaviour, in: *Revue Européenne des Sciences Sociales* 26, S. 19-41.
- Ströker, E. (1992), *Einführung in die Wissenschaftstheorie*, 4. Aufl., Darmstadt.
- Stuart, A./Kendall, M. (Hgg.) (1971), *The statistical papers of George Udny Yule*, New York.
- Student (1914-1915), The elimination of spurious correlation due to position in time or space, in: *Biometrika* 10, S. 179-180.

- Summers, L. H. (1991), The scientific illusion in empirical macroeconomics, in: *Scandinavian Journal of Economics* 93, S. 129-148. [wiederabgedruckt in Poirier (1994a)]
- Swamy, P. A./Conway, R. K./von zur Muehlen, P. (1985), The foundations of econometrics – Are there any? (with discussion), in: *Econometric Reviews* 4, S. 1-119.
- Swedberg, R. (1991), *Joseph A. Schumpeter. His life and work*, Cambridge.
- Swedberg, R. (1994), *Joseph A. Schumpeter. Eine Biographie*, Stuttgart.
- Tewes, T. (1991), Morgan, Mary S., The history of econometric ideas: Cambridge u. a., Cambridge Univ. Press, 1990, in: *Weltwirtschaftliches Archiv* 127, S. 614-615.
- Thome, H. (1988), Probleme des Identifizierens und Testens von Kausalbeziehungen in der statistischen Zeitreihenanalyse, in: Meier, F. (Hg.), *Prozeßforschung in den Sozialwissenschaften*, Stuttgart, S. 93-117.
- Thome, H. (1989), *Grundkurs Statistik für Historiker. Teil I: Deskriptive Statistik*, Köln (= Historical Social Research/Historische Sozialforschung Supplement 2).
- Thome, H. (1990), *Grundkurs Statistik für Historiker. Teil II: Induktive Statistik und Regressionsanalyse*, Köln (= Historical Social Research/Historische Sozialforschung Supplement 3).
- Thome, H. (1991), *Zeitreihenanalytische Verfahren für die Sozialwissenschaften*, Köln (Skript).
- Thome, H. (1992a), Beschreibende Zeitreihenanalyse: Komponentenzerlegung mit Hilfe gleitender Mittelwerte, in: *Historical Social Research/Historische Sozialforschung* 17/3, S. 63-105.
- Thome, H. (1992b), Review [zu Jarausch/Hardy (1991)], in: *Historical Social Research/Historische Sozialforschung* 17/2, S. 121-123.
- Thome, H. (1994a), Univariate Box/Jenkins-Modelle in der Zeitreihenanalyse, in: *Historical Social Research/Historische Sozialforschung* 19/3, S. 5-77.
- Thome, H. (1994b), *Soziologische Zyklentheorien zwischen Spekulation und Empirie*, Halle (= Der Hallesche Graureiher 94-5. Forschungsberichte des Instituts für Soziologie).
- Thome, H. (1994c), Berücksichtigung instabiler Varianzen in der Zeitreihenanalyse, in: *ZA-Information* 35, S. 82-109.
- Thome, H. (1996), Trends, cycles, and co-integration. Some issues in modelling long-term development in time series analysis, in: *Historical Social Research/Historische Sozialforschung* 21/4, S. 3-23.
- Thome, H. mit Rahlf, T. (1996), Dubious cycles: a methodological critique of the Namenwirth-Weber thesis on cultural change with an introduction into filter design methods, in: *Quality and Quantity* 30, S. 427-448.
- Tiao, G. C. (Hg.) (1984), *The collected works of George E. P. Box*, Belmont.
- Tilly, R. (1996), Wirtschaftsgeschichte als Disziplin, in: Ambrosius, G. et al. (Hgg.), *Moderne Wirtschaftsgeschichte. Eine Einführung für Historiker und Ökonomen*, München/Wien, S. 11-26.
- Tilly, R./Wellenreuther, T. (1985), Bevölkerungswanderung und Wohnungsbauzyklen in deutschen Großstädten im 19. Jahrhundert, in: Teuteberg, H. J. (Hg.), *Homo Habitans. Zur Sozialgeschichte des ländlichen und städtischen Wohnens in der Neuzeit*, Münster, S. 273-300.

- Tinbergen, J. (1935), Annual survey: suggestions on quantitative business cycle theory, in: *Econometrica* 3, S. 241-308.
- Tinbergen, J. (1937), *An econometric approach to business cycle problems*, Paris. [wiederabgedruckt in Wallis (1994a)]
- Tinbergen, J. (1939a), *Statistical testing of business-cycle theories. Bd. 1. A method and its application to investment activity*, Genf. [teilweise wiederabgedruckt in Hendry/Morgan (1995a)]
- Tinbergen, J. (1939b), *Statistical testing of business-cycle theories. Bd. 2. Business cycles in the United States of America, 1919-1932*, Genf. [teilweise wiederabgedruckt in Hendry/Morgan (1995a)]
- Tinbergen, J. (1940a), A reply [zu Keynes (1939a)], in: *Economic Journal* 50, S. 141-156. [wiederabgedruckt in Darnell (1994a) und Poirier (1994a)]
- Tinbergen, J. (1940b), Econometric business cycle research, in: *Review of Economic Studies* 7, S. 73-90. [wiederabgedruckt in Darnell (1994a) und Hendry/Morgan (1995a)]
- Tinbergen, J. (1952), *Einführung in die Ökonometrie*, Wien/Stuttgart.
- Tinbergen, J. (1963), *Lessons from the past*, Amsterdam u. a.
- Tintner, G. (1938), A note on economic aspects of the theory of errors in time series, in: *Quarterly Journal of Economics* 53, S. 141-149.
- Tintner, G. (1949), Foundations of probability and statistical inference (with discussion), in: *Journal of the Royal Statistical Society A* 112, S. 251-286. [wiederabgedruckt in Darnell (1994a) (ohne Diskussion)]
- Tintner, G. (1960), *Ökonometrie*, Göttingen.
- Toepel, A. (Hg.) (1994), *Sébastien le Pestre de Vauban, Projekt eines königlichen Zehnten [1707]*, Berlin.
- Tooke, T./Newmarch, W. (1858-1859), *Die Geschichte und Bestimmung der Preise während der Jahre 1793-1857. 2 Bde.*, Dresden.
- Trader, R. L. (1982), Bayes, Thomas, in: Kotz, S./Johnson, N. L. (Hgg.), *Encyclopedia of statistical sciences. Suppl.*, New York u. a., S. 14-17.
- Tsoulouhas, T. (1992), A new look at demographic and technological changes: England, 1550 to 1839, in: *Explorations in Economic History* 29, S. 169-203.
- Tufte, E. R. (1983), *The visual display of quantitative information*, Cheshire.
- Tukey, J. W. (1962), The future of data analysis, in: *Annals of Mathematical Statistics* 33, S. 1-67. [wiederabgedruckt in Jones (1986), teilweise wiederabgedruckt in Kotz/Johnson (1992a)]
- Tukey, J. W. (1968), An introduction to the calculations of numerical spectrum analysis, in: Harris, B. (Hg.), *Spectral analysis of time series*, New York, S. 25-46. [wiederabgedruckt in Brillinger (1984)]
- Tukey, J. W. (1972), Data analysis, computation and mathematics, in: *Quarterly of Applied Mathematics* 30, S. 51-65. [wiederabgedruckt in Jones (1986)]
- Tukey, J. W. (1977), *Exploratory data analysis*, Reading, Mass.
- Tukey, J. W. (1980a), We need both exploratory and confirmatory, in: *American Statistician* 34, S. 23-25. [wiederabgedruckt in Jones (1986)]
- Tukey, J. W. (1980b), Can we predict where 'time series' should go next?, in: Brillinger, D. R./Tiao, T. C. (Hgg.), *Directions in time series*, Hayward, S. 1-31. [wiederabgedruckt in Brillinger (1984)]

- Tukey, J. W. (1984), Foreword to the time series volumes, in: Brillinger, D. R. (Hg.), *The collected works of John W. Tukey. Bd. 1. Time series: 1949-1964. Bd. 2. Time series: 1965-1984*, Belmont, S. xxxvi-xlii.
- Tukey, J. W. (1986a), Foreword to the philosophy volumes, in: Jones, L. V. (Hg.), *The collected works of John W. Tukey. Bd. 3. Philosophy and principles of data analysis: 1949-1964*, Belmont, S. xxxix-xliv.
- Tukey, J. W. (1986b), Data analysis and behavioral science or learning to bear the quantitative man's burden by shunning badmandments (1961), in: Jones, L. V. (Hg.), *The collected works of John W. Tukey. Bd. 3. Philosophy and principles of data analysis: 1949-1964*, Belmont, S. 187-389.
- Usher, A. P. (1932), The application of the quantitative method to economic history, in: *Journal of Political Economy* 40, S. 186-209.
- Usher, A. P. (1949), The significance of modern empiricism for history and economics, in: *Journal of Economic History* 9, S. 131-155.
- Vauban, S. (1994), *Projekt eines königlichen Zehnten [1707]. Übersetzt und herausgegeben von A. Toepel*, Berlin.
- Vining, R. (1949a), Koopmans on the choice of variables to be studied and of methods of measurement, in: *Review of Economics and Statistics* 31, S. 77-86. [wiederabgedruckt in Darnell (1994a), Poirier (1994a) und Hendry/Morgan (1995a)]
- Vining, R. (1949b), A rejoinder [to Koopmans (1949)], in: *Review of Economics and Statistics* 31, S. 91-94. [wiederabgedruckt in Darnell (1994a), Poirier (1994a) und Hendry/Morgan (1995a)]
- von Below, G. (1893), Rezension: Deutsche Geschichte von Karl Lamprecht. I-III, in: *Historische Zeitschrift* 71, S. 465-498.
- von Below, G. (1920), *Probleme der Wirtschaftsgeschichte. Eine Einführung in das Studium der Wirtschaftsgeschichte*, Tübingen.
- von Huhn, R. (1931), A trigonometrical method for computing the scales of statistical charts to improve visualization, in: *Journal of the American Statistical Association* 26, S. 319-324.
- von Mises, R. (1928), *Wahrscheinlichkeit, Statistik und Wahrheit*, Wien.
- von Randow, G. (1992), *Das Ziegenproblem. Denken in Wahrscheinlichkeiten*, Reinbek.
- Wald, A. (1950), *Statistical decision functions*, New York.
- Wald, A. (1936), *Berechnung und Ausschaltung von Saisonschwankungen*, Wien.
- Walker, G. (1931), On periodicity in series of related terms, in: *Proceedings of the Royal Society A* 131, S. 518-532.
- Walker, H. M. (1929), *Studies in the history of statistical method*, Baltimore.
- Wallis, K. F. (1977), Multiple time series analysis and the final form of econometric models, in: *Econometrica* 45, S. 1481-1497. [wiederabgedruckt in Wallis (1994a)]
- Wallis, K. F. (Hg.) (1994a), *Macroeconometric modelling. Bd. 1. The first thirty years. Bd. 2. From the 1960s*, Aldershot (= The International Library of Critical Writings in Econometrics 2).
- Wallis, K. F. (1994b), Introduction, in: Wallis, K. F. (Hg.), *Macroeconometric modelling. Bd. 1. The first thirty years*, Aldershot (= The International Library of Critical Writings in Econometrics 2), S. ix-xxvi.
- Walter, R. (1994), *Einführung in die Wirtschafts- und Sozialgeschichte*, Paderborn.

- Watson, G. S. (1983), Hypothesis testing, in: Kotz, S./Johnson, N. L. (Hgg.), *Encyclopedia of statistical sciences*. Bd. 3, New York u. a., S. 712-722.
- Weber, M. (1903-1906), Roschers 'historische Methode'; Knies und das Irrationalitätsproblem, in: *Schmollers Jahrbuch* 27, 29 und 30. [unter dem Titel "Roscher und Knies und die logischen Probleme der historischen Nationalökonomie" wiederabgedruckt in Winckelmann (1968)]
- Weber, M. (1906), Zur Auseinandersetzung mit Eduard Meyer; Objektive Möglichkeit und adäquate Verursachung in der historischen Kausalbetrachtung, in: *Archiv für Sozialwissenschaft und Sozialpolitik* 22. [wiederabgedruckt unter dem Titel "Kritische Studien auf dem Gebiet der kulturwissenschaftlichen Logik" in Winckelmann (1968)]
- Wegman, E. J./Carr, D. B. (1993), Statistical graphics and visualization, in: Rao, C. R. (Hg.), *Handbook of statistics*. Bd. 9, New York, S. 857-958.
- Wehler, H.-U. (Hg.) (1973a), *Geschichte und Ökonomie*, Köln (= Neue wissenschaftliche Bibliothek 58).
- Wehler, H.-U. (1973b), Einleitung, in: Wehler, H.-U. (Hg.), *Geschichte und Ökonomie*, Köln (= Neue wissenschaftliche Bibliothek 58).
- Weiller, K./Mirowski, P. (1990), Rates of interest in 18th-century England, in: *Explorations in Economic History*, S. 1-28.
- Weiss, L. (1992), Introduction to Wald (1949) Statistical decision functions, in: Kotz, S./Johnson, N. L. (Hgg.), *Breakthroughs in statistics*. Bd. 1. *Foundations and basic theory*, New York u. a., S. 335-341.
- Westergaard, H. (1932), *Contributions to the history of statistics*, London.
- West, M./Harrison, J. (1989), *Bayesian forecasting and dynamic models*, New York u. a. (= Springer Series in Statistics).
- West, M./Harrison, P. J./Migon, H. (1985), Dynamic generalized linear models and Bayesian forecasting (with comments), in: *Journal of the American Statistical Association* 80, S. 73-97.
- Whaples, R. (1991), A quantitative history of the Journal of Economic History and the cliometric revolution, in: *Journal of Economic History* 51, S. 289-301.
- Wiener, N. (1930), Generalized harmonic analysis, in: *Acta Mathematicae* 55, S. 117-258.
- Wiener, N. (1962), *Mathematik: Mein Leben*, Düsseldorf/Wien.
- Wild, C. J. (1994), Embracing the 'wider view' of statistics, in: *American Statistician* 48, S. 163-171.
- Williamson, J. G. (1972), Embodiment, disembodiment, learning by doing, and returns to scale in nineteenth-century cotton textiles, in: *Journal of Economic History* 32, S. 691-705.
- Williamson, S. H. (1991), The history of cliometrics, in: Mokyr, J. (Hg.), *The vital one. Essays in honor of Jonathan R. T. Hughes*, Greenwich (= Research in Economic History, Supplement 6), S. 15-31.
- Winckelmann, J. (Hg.) (1968), *Gesammelte Aufsätze zur Wissenschaftslehre von Max Weber*, 3. Aufl., Tübingen.
- Wise, M. N. (1987), How do sums count? On the cultural origins of statistical causality, in: Krüger, L./Daston, L. J./Heidelberger, M. (Hgg.), *The probabilistic revolution*. Bd. 1. *Ideas in history*, Cambridge u. a., S. 395-425.

- Wold, H. O. (1938), *A study in the analysis of stationary time series*, Uppsala.
- Wold, H. O. (1956), Causal inference from observational data: a review of ends and means, in: *Journal of the Royal Statistical Society A* 119, S. 28-60.
- Wold, H. O. (Hg.) (1965a), *Bibliography on time series and stochastic processes. An international team project*, Edinburgh/London.
- Wold, H. O. (1965b), A graphic introduction to stochastic processes, in: Wold, H. O. (Hg.), *Bibliography on time series and stochastic processes*, Edinburgh/London, S. 7-61.
- Wold, H. O. (1969), Econometrics as pioneering in nonexperimental model building, in: *Econometrica* 37, S. 369-381. [wiederabgedruckt in Darnell (1994a) und Poirier (1994a)]
- Wolters, J. (1987), Ökonometrische Modelle bei Zeitreihendaten versus multivariate Zeitreihenmodelle – eine Übersicht, in: *Statistische Hefte* 28, S. 1-25.
- Wonnacott, R. J./Wonnacott, T. H. (1990), *Introductory statistics*, 6. Aufl., New York.
- Working, H. (1922), The determination of secular trend reconsidered, in: *Journal of the American Statistical Association* 18, S. 497-502.
- Working, H. (1934), A random-difference series for the use in the analysis of time series, in: *Journal of the American Statistical Association* 29, S. 11-24. [teilweise wiederabgedruckt in Hendry/Morgan (1995a)]
- Wright, C. W. (1938), The nature and objectives of economic history, in: *Journal of Political Economy* 46, S. 688-701.
- Yule, G. U. (1895), On the correlation of total pauperism with proportion of out-relief, I: all ages, in: *Economic Journal* 5, S. 603-611.
- Yule, G. U. (1896a), Notes on the history of pauperism in England and Wales from 1850, treated by the method of frequency-curves; with an introduction on the method, in: *Journal of the Royal Statistical Society* 59, S. 318-357.
- Yule, G. U. (1896b), On the correlation of total pauperism with proportion of out-relief, II: males over sixty-five, in: *Economic Journal* 6, S. 613-623.
- Yule, G. U. (1897), On the theory of correlation, in: *Journal of the Royal Statistical Society* 60, S. 812-854.
- Yule, G. U. (1906), On the changes in marriage- and birth-rates in England and Wales during the past half century; with an inquiry as to their probable causes (with discussion), in: *Journal of the Royal Statistical Society* 69, S. 88-147.
- Yule, G. U. (1915), Review of Moore, Economic cycles, in: *Journal of the Royal Statistical Society* 78, S. 302-305.
- Yule, G. U. (1921), Fall of the birth rate, in: *Journal of the Royal Statistical Society* 84, S. 128.
- Yule, G. U. (1926), Why do we sometimes get nonsense correlations between time-series? – A study in sampling and the nature of time-series (with discussion), in: *Journal of the Royal Statistical Society* 89, S. 1-69. [wiederabgedruckt (ohne Diskussion) in Stuart/Kendall (1971) und Darnell (1994a), teilweise wiederabgedruckt in Hendry/Morgan (1995a)]
- Yule, G. U. (1927), On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers, in: *Philosophical Transactions of the Royal Society of London A* 226, S. 267-298. [wiederabgedruckt in Stuart/Kendall (1971), teilweise wiederabgedruckt in Hendry/Morgan (1995a)]

- Yule, G. U. (1942), Discussion [zu Kendall, The future of statistics], in: *Journal of the Royal Statistical Society* 105, S. 83-84.
- Zabell, S. L. (1989), R. A. Fisher on the history of inverse probability, in: *Statistical Science* 4, S. 247-263.
- Zadeh, L. (1965), Fuzzy sets, in: *Information and Control* 8, S. 338-353.
- Zellner, A. (1971), *An introduction to Bayesian statistics in econometrics*, New York.
- Zellner, A. (Hg.) (1980), *Bayesian analysis in econometrics and statistics. Essays in honor of Harold Jeffreys*, Amsterdam (= Studies in Bayesian Econometrics 1).
- Zellner, A. (1985), Bayesian econometrics, in: *Econometria* 53, S. 253-269. [wiederabgedruckt in Polson/Tiao (1995a)]
- Zellner, A. (1988), Bayesian analysis in econometrics, in: *Journal of Econometrics* 37, S. 27-50. [wiederabgedruckt in Granger (1990a)]
- Zellner, A. (1990), Bayesian inference, in: Eatwell, J./Milgate, M./Newman, P. (Hgg.), *The new Palgrave: time series and statistics*, New York/London, S. 36-61.
- Zellner, A. (1992), Review [von Morgan (1990)], in: *Journal of Political Economy* 100, S. 218-222.
- Zellner, A./Palm, F. (1974), Time series analysis and simultaneous equation econometric models, in: *Journal of Econometrics* 2, S. 17-54. [wiederabgedruckt in Harvey (1994a)]
- Zorn, W. (1974), Das Fach Wirtschafts- und Sozialgeschichte im letzten halben Jahrhundert, in: Bog, I. et al. (Hgg.), *Wirtschaftliche und soziale Strukturen im säkularen Wandel. Festschrift für Wilhelm Abel zum 70. Geburtstag. Bd. 1*, Hannover, S. 11-22.
- Zorn, W. (1982), Wirtschaftsgeschichte, in: Albers, W. et al. (Hgg.), *Handwörterbuch der Wirtschaftswissenschaft. Bd. 9*, Stuttgart u. a., S. 55-83.
- Zorn, W. (1991), Hermann Kellenbenz, 1913-1990, in: *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte* 78, S. 1-5.

Wirtschafts- geschichte

Ökonometrie

Statistik

1840er: Ältere historische Schule

1883/1884: Methodenstreit Menger / Schmoller

1838/1857: Tooke / Newmarch, Price History

1854: Buckle, History of Civilization

1866ff: Rogers, Price History

1893: Ashley, Study of Ec. His

1862: Juglarzyklen

1884: Jevons Sonnenzyklentheorien

1835: Quetelets Physique sociale

1888: Galton führt Co-relation ein

1883/1885 Edgeworth über theoreti. Statistik

